

UNIVERSIDAD DE CUENCA



FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS

Buscador semántico universitario: Caso de estudio Universidad de Cuenca

Tesis previa a la obtención del Título de Ingeniero en Sistemas

Autores:

Yolanda del Carmen Aucapiña Pineda C.I. 0105290456

Carlos Orlando Plaza Villa C.I. 0107064180

Director:

Ing. Pablo Fernando Vanegas Peralta, PhD. C.I. 0102274891

Co-director:

Ing. Víctor Hugo Saquicela Galarza, PhD. C.I. 0103599577

Cuenca - Ecuador

2018

Resumen

El uso de tecnología semántica provee de características como: notación semántica, generación de nuevo conocimiento y modelo de datos interoperables a los recursos de la web, por medio de modelos ontológicos comunes que permiten enlazar datos y no páginas web como en la web tradicional, permitiendo al ordenador ser capaz de entender los datos y retornar información precisa a los usuarios.

En la Universidad de Cuenca se generan grandes volúmenes de datos que han sido preservados a lo largo del tiempo; por lo tanto, la información se ha almacenado en diferentes tecnologías del momento, lo cual ha dado lugar a información universitaria desintegrada tanto a nivel de modelos de datos como tecnologías de almacenamiento (fuentes de datos heterogéneas). En este contexto, el proceso de recuperar información de dichas fuentes requiere un tiempo considerable debido a la necesidad de realizar procesos complejos para integrar y extraer únicamente la información útil para el usuario.

Como solución a la recuperación de información, se ha utilizado tecnología semántica con el propósito de aprovechar sus principales ventajas. De esta manera, se desarrolla un modelo ontológico que representa el dominio universitario, permitiendo integrar los datos procedentes de fuentes heterogéneas, para posteriormente, mediante el prototipo de un único buscador que trabaja sobre el modelo de integración, extraer y visualizar información a través de simples consultas realizadas por el usuario. Además, las diversas etapas llevadas a cabo para la consecución del prototipo del buscador semántico siguen metodologías probadas, y en ciertos casos son soportados por procesos automatizados.

Palabras clave: Buscador semántico, Datos enlazados, ETL, Integración, NeOn, Ontología, RDF, Web semántica

Abstract

Semantic technology provides characteristics such as: semantic notation, generation of new knowledge and interoperable data model to resources on web. Through common ontological models that allow linking data and not web pages as in traditional web, by allowing computer to be able to understand the data and return accurate information to users.

In University of Cuenca large volumes of data are generated and they have been preserved over time; therefore, the information has been stored in different technologies at the moment, which has led to disintegrated university information both at the level of data models and storage technologies (heterogeneous data sources). In this context, the process to retrieve information from these sources requires considerable time due to the need to perform complex processes to integrate and extract only information useful to the user.

As solution to the recovery of information, semantic technology has been used with the purpose of taking advantage of its main features. In this way, an ontological model is developed that represents the university domain, by allowing to integrate data from heterogeneous sources, for later, through the prototype of a single search engine that works on the integration model, extract and visualize information through simple queries write by the user. In addition, several stages carried out for the achievement of the semantic search engine prototype follow proven methodologies, and in some cases are supported by automated processes.

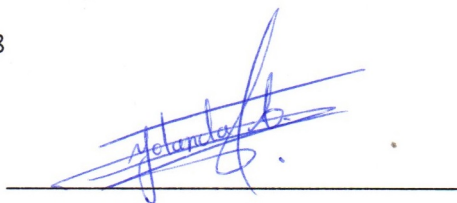
Keywords: ETL, Integration, Linked data, NeOn, Ontology, RDF, Search engine, Semantic web

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Yolanda del Carmen Aucapiña Pineda en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación "Buscador semántico universitario: Caso de estudio Universidad de Cuenca", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 20 de Marzo de 2018



Yolanda del Carmen Aucapiña Pineda

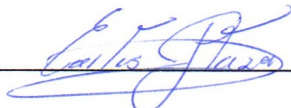
C.I: 0105290456

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Carlos Orlando Plaza Villa en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación “Buscador semántico universitario: Caso de estudio Universidad de Cuenca”, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 20 de Marzo de 2018



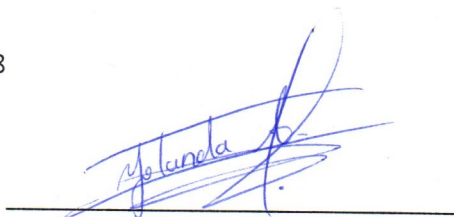
Carlos Orlando Plaza Villa

C.I: 0107064180

Cláusula de Propiedad Intelectual

Yolanda del Carmen Aucapiña Pineda, autor/a del trabajo de titulación “Buscador semántico universitario: Caso de estudio Universidad de Cuenca”, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 20 de Marzo de 2018



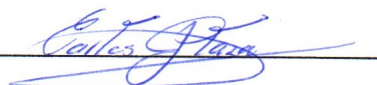
Yolanda del Carmen Aucapiña Pineda

C.I: 0105290456

Cláusula de Propiedad Intelectual

Carlos Orlando Plaza Villa, autor/a del trabajo de titulación “Buscador semántico universitario: Caso de estudio Universidad de Cuenca”, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 20 de Marzo de 2018



Carlos Orlando Plaza Villa

C.I: 0107064180

Índice general

Resumen	1
Lista de figuras	10
Lista de tablas	13
Dedicatoria	14
Agradecimientos	15
1. Introducción	17
1.1. Planteamiento del problema y justificación	17
1.2. Objetivos del trabajo	19
1.2.1. Objetivo general	19
1.2.2. Objetivos específicos	19
1.3. Metodología de investigación	19
1.4. Estructura de la tesis	20
2. Fundamentos teóricos	22
2.1. Evolución de la web	22
2.2. Web Semántica	23
2.2.1. Arquitectura	23
2.3. Ontologías	25
2.3.1. Elementos básicos de una ontología	26
2.4. Linked Open Data	27
2.5. Vocabularios ontológicos	28
2.6. Tecnologías semánticas	29
2.6.1. RDF (Resource Description Framework)	29
2.6.2. RDF – Schema (RDFS)	30
2.6.3. SPARQL (Simple Protocol and RDF Query Language)	31
2.6.4. OWL (Web Ontology Language)	32
2.7. Ciclos de vida para datos enlazados	33

2.7.1.	Ciclo de vida de Hyland	33
2.7.2.	Ciclo de vida de Hausenblas	34
2.7.3.	Ciclo de vida de Villazón - Terrazas	35
2.7.4.	Selección de ciclo vida a utilizar	37
2.8.	Metodología NeOn para el desarrollo ontológico	39
2.8.1.	El marco de trabajo de la metodología NeOn	39
2.8.2.	Escenarios para construir redes ontológicas	39
2.8.3.	Actividades de la ingeniería ontológica	45
2.8.4.	Modelos de ciclo de vida para redes ontológicas	52
3.	Trabajos relacionados	55
3.1.	Introducción	55
3.2.	Modelos propuestos	55
3.2.1.	Modelo ontológico para representar y publicar datos académicos . .	56
3.2.2.	Linked Data Academic University (LDAU)	57
3.2.3.	Ontology for Linked Open University Data (OLOUD)	59
3.2.4.	UH-Ontology	60
3.2.5.	Higher Education Reference Ontology (HERO)	62
3.3.	Resultados obtenidos	64
4.	Proceso de integración de datos	67
4.1.	Especificación	68
4.1.1.	Definición del alcance total	68
4.1.2.	Identificación y análisis de fuentes de datos	71
4.1.3.	Diseño de URI	73
4.2.	Modelo ontológico (NeOn)	73
4.2.1.	Especificación de requerimientos	74
4.2.2.	Selección del ciclo de vida y escenarios	76
4.2.3.	Reutilización ontologías de dominio	79
4.2.4.	Fusión de recursos ontológicos	84
4.2.5.	Re-estructuración de recursos ontológicos	85
4.2.6.	Implementación	88
4.3.	Generación RDF	91
4.4.	Publicación	98
4.5.	Explotación	101
4.6.	Resultados	103
5.	Prototipo del buscador	111
5.1.	Descripción general	111
5.2.	Funcionamiento	114
5.3.	Presentación de resultados	115

6. Conclusiones y trabajos futuros	118
6.1. Conclusiones finales	118
6.2. Líneas de trabajos futuros	119
A. Documento de especificación de requerimientos ontológicos	120
B. Vocabularios reutilizados en la red ontológica	132
C. Diagrama de la red ontológica	134
D. Clases definidas en el vocabulario UCuenca-Ontology	138
E. Propiedades de objetos definidas en el vocabulario UCuenca-Ontology	142
F. Propiedades de datos definidas en el vocabulario UCuenca-Ontology	145
G. Descripción de clases definidas en la red ontológica UCuenca-Ontology	150
H. SQL para extracción de datos	154
I. Problemas y soluciones	190
I.1. Extracción de datos	190
I.2. Tranformaciones	190
I.2.1. Ampliación de memoria	190
I.2.2. Almacenamiento de registros	191
I.2.3. Plugin Data Precatching	192
I.2.4. Error en la base de datos interna	193
I.2.5. Error en generación Elda	193
I.3. Buscador semántico	195
I.3.1. Visualización gráfica	195
Bibliografía	196

Índice de figuras

2.1. Arquitectura de la web semántica[1].	24
2.2. Ejemplo ilustrativo de los elementos de una ontología	26
2.3. Estructura de una tripleta RDF	30
2.4. Ejemplo de consulta en SPARQL	32
2.5. Actividades propuestas en el ciclo de vida de Hyland[2]	33
2.6. Actividades establecidas en el ciclo de vida de Hausenblas [2]	34
2.7. Actividades definidas en el ciclo de vida de Villazón – Terrazas[2]	35
2.8. Ficha descriptiva de la actividad de especificación de requerimientos ontológicos [3]	46
2.9. Plantilla para el ORSD [3]	48
2.10. Ciclo de vida en cascada para redes ontológicas[4]	52
2.11. Esquema del modelo de ciclo de vida iterativo incremental[4]	53
3.1. Modelo conceptual de la ontología de planes de curso	57
3.2. Clases y relaciones definidas en la ontología LDAU.	58
3.3. Clases y propiedades del módulo principal de OLOUD.	60
3.4. Estructura del módulo de localización/navegación en interiores.	61
3.5. Clase Agent y sus relaciones con otras clases y subclases	62
3.6. Preglosario de términos junto con sus respectivas frecuencias.	63
3.7. Jerarquía de conceptos clave de HERO	64
4.1. Organigrama que describe el alcance académico del proyecto.	71
4.2. Metamodelo de la red ontológica obtenida al fusionar las ontologías seleccionadas	85
4.3. Diagrama general de la red ontológica UCuenca-Ontology	87
4.4. Anotaciones y parte de los vocabularios importados a la red ontológica	88
4.5. Modelado de clases de la red ontológica	89
4.6. Relaciones establecidas para la clase Student	90
4.7. Anotaciones incluidas en el concepto "AcademicOffer"	90
4.8. Proceso ETL para la generación RDF	91
4.9. Definición del paso para extraer información de las bases de datos Oracle	92
4.10. Lectura de datos de una fuente de extensión .csv	93

4.11. Casting de datos para homogeneizar la información	93
4.12. Unión de datos provenientes de fuentes heterogéneas	94
4.13. Conexión a bases de datos embebida para el almacenamiento temporal de datos	94
4.14. Lectura del modelo ontológico desde archivo con extensión OWL	95
4.15. Bloqueo de flujo debido a dependencias entre pasos en PDI	95
4.16. Reglas de mapeo entre datos y clases	96
4.17. Reglas de mapeo entre datos y propiedades de datos	97
4.18. Reglas de mapeo entre datos y propiedades de objetos	97
4.19. Configuración del componente RDF Generation	98
4.20. Proyecto Fuseki	99
4.21. Presentación de contenido - Servicio Elda	100
4.22. Fuseki Endpoint	101
4.23. Configuración del plugin para la generación de un proyecto Elda	102
4.24. Presentación de contenido - Servicio Elda	103
4.25. Extracto RDF - Información Universidad de Cuenca	104
4.26. Condición de equivalencia definida en la clase Student	105
4.27. SPARQL - Personas	106
4.28. SPARQL - Estudiantes	107
4.29. SPARQL - Profesores	108
4.30. SPARQL - Jefes Familia	108
4.31. SPARQL - Estudiante y Jefe Familia	109
4.32. SPARQL - Estudiantes de carrera X	109
4.33. SPARQL - Carrera X de estudiantes	110
5.1. Proyecto - buscador Fedquest	112
5.2. Registro Endpoint	112
5.3. Definición y configuración de entidades	113
5.4. Selección de entidades - vista en modo grafo	114
5.5. Selección entidades - filtros	114
5.6. Buscador semántico: Ventana Inicial	115
5.7. Resultados tras buscar "Facultad de Ingeniería"	115
5.8. Visualización de información relacionada a "Facultad de Ingeniería"	116
5.9. Resultados desplegados en modo grafo - Facultad de Ingeniería	117
5.10. Expansión del grafo - Facultad de Ingeniería	117
C.1. Diagrama de la red ontológica: Parte 1	135
C.2. Diagrama de la red ontológica: Parte 2	136
C.3. Diagrama de la red ontológica: Parte 3	137
I.1. Incremento de memoria a PDI	191

I.2.	Incremento de memoria a PDI	192
I.3.	Corrección de transformación	192
I.4.	Error en base de datos interna	193
I.5.	Corrección de entidades - Elda	194
I.6.	Definición de referencias a recursos - Elda	194
I.7.	Corrección de vista gráfica	195

Índice de tablas

2.1. Ejemplo de representación de una tripleta mediante URIs	30
2.2. Clases y propiedades contenidas en RDF-Schema[5].	31
2.3. Alcance de actividades definidas en los diferentes ciclos de vida[6]	38
2.4. Plantilla que contiene los criterios de decisión para seleccionar ontologías de dominio. [7]	51
3.1. Resumen de trabajos relacionados	65
4.1. Fuentes de datos obtenidas como resultado del proceso de identificación y análisis	71
4.2. Diseño de URIs a utilizar en el proyecto de datos enlazados	73
4.3. Documento de especificación de requerimientos: Secciones 1 a 5	75
4.4. Parte del grupo de preguntas de competencia sobre organización académica	76
4.5. Parte de los conceptos a incluir en el modelo ontológico	77
4.6. Preguntas para establecer ciclo de vida y escenarios a utilizar	78
4.7. Ontologías candidatas por grupo	80
4.8. Propósito y alcance de ontologías candidatas para los grupos organización administrativa y organización académica	81
4.9. Resultado de la valoración de ontologías candidatas para el grupo organi- zación administrativa	82
4.10. Resultado de la valoración de ontologías candidatas para el grupo organi- zación académica	82
4.11. Selección de ontologías para la organización administrativa	83
4.12. Ontologías seleccionadas para ser reutilizadas	84

Dedicatoria

A la persona más especial e importante, porque desde que Dios te envió a mi vida te convertiste en mi razón para levantarme cada día y esforzarme por un futuro mejor, a ti hijo mío, porque con tu sonrisa, tus abrazos y tu cariño haces que todo valga la pena. Recuerda que te adoro y que siempre, sin importar las circunstancias, podrás contar conmigo.

Yolanda Aucapiña

Este trabajo es dedicado a mis padres Laura y Samuel, y a mis hermanos quienes me brindaron fortaleza y apoyo durante mi formación personal y académica. Por ayudarme afrontar los retos, las situaciones difíciles; y por prepararme para la vida a través de sus consejos y enseñanzas un especial agradecimiento para ti papá y mamá, gracias por estar junto a mi.

Carlos Plaza

Agradecimientos

Con un inmenso sentimiento de cariño y respeto hago uso de este espacio para expresar mi gratitud a todas aquellas personas que de una u otra manera fueron un apoyo invaluable para culminar mi carrera universitaria de forma exitosa, porque estoy segura que sin ustedes nada de esto hubiese sido posible.

Primero, gracias a Dios por su infinito amor y bondad, por estar a mi lado en todo momento, por permitirme ver cumplida una de mis metas y por poner en el camino personas tan maravillosas.

A mis padres Juan y Leticia, por ser el pilar fundamental de todo lo que soy, gracias por dedicar su tiempo y esfuerzo para hacer de mi una persona de bien. A mis herman@s Isabel, Ángel y Hernán, gracias por ser ejemplo y guía, por sus palabras de aliento, porque cada uno a su manera estuvo a mi lado brindándome su apoyo. ¡Son los mejores hermanos míos! A mis cuñad@s Manuel, Fanny y Nancy, gracias por sus consejos, compañía y amistad.

A mi adorado hijo, Fernando, por ser mi mayor motivación, gracias por alegrar mi vida con tu sonrisa, porque con tu afecto y cariño logras que cada día sea mucho más verdadero. Eres un niño muy especial, para mi el mejor de todos, me enorgullece ser tu mamá.

A los docentes que durante estos años de formación universitaria tuve el gusto de conocer, a los compañeros que a lo largo de los años se convirtieron en grandes amigos y a todas aquellas personas que de una forma u otra fueron parte de este camino.

Yolanda Aucapiña

Agradecimientos

Por luchar por mí y por mis hermanos un total y sincero agradecimiento a mis padres Laura y Samuel, quienes han mostrado siempre su amor incondicional guiándonos y apoyándonos en cualquier momento y circunstancia de la vida, porque sin importar cuán adverso este sea siempre han estado ahí.

A mis hermanos, sobrinos y cuñados quienes siempre me brindaron su apoyo absoluto, alentándome a seguir adelante y no decaer jamás, gracias por estar junto a mí en los momentos difíciles.

A la Universidad de Cuenca por formar profesionales con capacidad autodidacta y a los docentes quienes compartieron y promovieron el interés por nuevo conocimiento basado en principios éticos para mejorar el mundo; ayer, hoy y mañana.

A los directores, y demás personas que participaron en este proyecto por apoyar con los recursos y el conocimiento necesario para llevar a cabo los objetivos planteados en este trabajo de titulación.

Carlos Plaza

Capítulo 1

Introducción

La web semántica plantea un nuevo enfoque en el que se proponen nuevas técnicas y paradigmas para la representación de la información y el conocimiento, esto facilita tanto la localización como el compartir, integrar y recuperar recursos. Dicho enfoque ha abierto un campo de investigación que propone enriquecer la estructura de la información y agregar componentes semánticos que puedan ser procesados de forma automática[8].

Considerando la importancia de agregar significado a la información existente y con el objetivo de solucionar la problemática que genera el disponer de numerosas fuentes de datos que a través del tiempo han hecho uso de diversas tecnologías, en el presente proyecto de investigación se describe el proceso realizado para integrar dichas fuentes de datos existentes en un dominio universitario, considerando como caso de uso inicial a la Universidad de Cuenca. Para esto es necesario llevar a cabo las diferentes actividades incluidas en el proceso de desarrollo de un proyecto de datos enlazados.

Entre las actividades a realizar, se considera la creación de un modelo ontológico, denominado UCuenca-Ontology, mediante el cual se representa la organización administrativa y académica de una universidad, y que posteriormente es utilizado como base para la integración de datos provenientes de las diversas fuentes, en un único recurso, sobre el que, mediante la aplicación de un buscador semántico, se permitirá al usuario ejecutar búsquedas de forma eficiente y visualizar los resultados obtenidos de una forma sencilla y amigable.

1.1. Planteamiento del problema y justificación

En el ámbito universitario es habitual que se genere una gran cantidad de información (alumnos, asignaturas, docentes, facultades, mallas curriculares, personal administrativo, etc.), proveniente de diferentes áreas y departamentos, lo que dificulta su adecuada ges-

ción, esto sumado al constante avance tecnológico que ha provocado que las instituciones tengan que adaptarse a las tecnologías del momento, ha dado como resultado un almacenamiento de datos en diversos formatos y haciendo uso de varias tecnologías; es decir, se han generado fuentes de datos heterogéneas.

En este escenario, realizar una búsqueda de información sobre todas las fuentes de datos existentes se convierte en una tarea compleja y que requiere un tiempo considerable, debido a que los datos se encuentran aislados y en ciertos casos es imposible enlazar unos con otros. Es por esto, que en la mayoría de los casos es necesario construir consultas SQL¹ complejas (considerando que incluso pueden ser distributivas) y en otros casos utilizar herramientas complementarias que permitan realizar esta tarea. Adicionalmente, la información universitaria no se encuentra anotada semánticamente, por lo que, su procesamiento se realiza por contenido y no por el significado, haciendo que los resultados obtenidos contengan inconsistencias que tienen que ser depuradas por el usuario.

Como solución a esta problemática este proyecto plantea presentar y desarrollar un modelo ontológico que describa el dominio académico y organizacional de una universidad haciendo uso de vocabularios existentes y tomando como caso de uso inicial a la Universidad de Cuenca. Adicionalmente, se propone un prototipo de un buscador semántico que permita al usuario obtener información específica de los datos disponibles en la institución universitaria. Dicha solución estará formada por un buscador que actúa sobre una arquitectura de integración creada siguiendo la guía de buenas prácticas establecida por la W3C².

Es claro que la web semántica ha abierto un nuevo campo de investigación que busca extender su uso a diversas áreas con la finalidad de mejorar los servicios ofrecidos al usuario. En el ámbito de la educación universitaria existen diferentes estudios centrados especialmente en brindar soluciones en lo relacionado al área académica, pero son mínimos los esfuerzos para crear un modelo general que integre la información de todas las fuentes heterogéneas disponibles en un único modelo. En base a esto, surgen las siguientes hipótesis de investigación:

1. La implementación de un modelo en una única tecnología semántica presentará una mejora considerable en la gestión de la información respecto a los modelos existentes en las bases de datos heterogéneas, dotando al mismo tiempo de un modelo de interoperabilidad.

¹Structured Query Language: Lenguaje de consulta estándar para bases de datos relacionales.

²World Wide Web Consortium <http://www.w3c.es/>: Consorcio internacional cuyo objetivo es generar recomendaciones acerca de los estándares de la World Wide Web (www).

2. El modelo ontológico creado para representar el ámbito organizacional y académico de la Universidad de Cuenca haciendo uso de metodologías y tecnologías del campo semántico, permitirá obtener resultados con mayor precisión y potencia ante una búsqueda, donde la información será comprendida tanto por el ordenador como por el usuario.

1.2. Objetivos del trabajo

Con la finalidad de guiar el desarrollo del presente proyecto de investigación se plantean los siguientes objetivos, los cuales a su vez, serán de utilidad para medir el nivel de cumplimiento al finalizar el proyecto.

1.2.1. Objetivo general

Definir una plataforma de integración de fuentes de datos utilizando tecnologías semánticas que permita al usuario recuperar información de forma ágil y eficiente de entre las grandes cantidades de datos disponibles en una universidad, tomando como caso de uso inicial a la Universidad de Cuenca.

1.2.2. Objetivos específicos

1. Conocer los conceptos de la web semántica, así como las metodologías existentes para el desarrollo de un proyecto de datos enlazados y seleccionar la que mejor se adapte al presente caso de estudio.
2. Conocer los trabajos realizados referentes a la aplicación de tecnología semántica en el campo universitario.
3. Aplicar la metodología seleccionada para diseñar un modelo que integre los datos contenidos en las diversas fuentes de datos existentes en la Universidad de Cuenca.
4. Crear un prototipo de buscador semántico que permita realizar búsquedas de datos sobre el modelo de integración.

1.3. Metodología de investigación

Las etapas a seguir para el desarrollo del presente proyecto se detallan a continuación:

1. **Búsqueda y clasificación bibliográfica:** Inicialmente y con el objetivo de obtener información que sirva como punto de partida para el desarrollo del proyecto, se plantea llevar a cabo una búsqueda de bibliografía relacionada al dominio, la cual posteriormente será clasificada descartando la información considerada irrelevante.

2. **Investigación bibliográfica:** Utilizando como base el conjunto de recursos bibliográficos seleccionados en la etapa anterior, se extraerá la información que pueda ser aplicada en el desarrollo del presente proyecto.
3. **Definición de metodologías a aplicar:** En base a la información obtenida hasta este punto se procede a seleccionar la metodología que permita cumplir con los objetivos del proyecto de la mejor manera, teniendo en consideración las guías de buenas prácticas para datos enlazados provistas por la W3C.
4. **Definición de las herramientas a utilizar:** Una vez establecida la metodología a aplicar, se requiere determinar las herramientas adecuadas que faciliten su ejecución, para ello, se realizará un análisis de las herramientas disponibles para la publicación de datos enlazados con la finalidad de definir la que mejor se acople a este caso.
5. **Diseño del modelo de integración:** Esta etapa implica inicialmente realizar un levantamiento de requerimientos, en base a los cuales, se llevará a cabo el diseño del modelo ontológico que permita integrar los datos disponibles en las diversas fuentes de datos.
6. **Extracción y Limpieza de Datos:** Una vez finalizada la etapa de modelado, se procederá con la extracción y limpieza de los datos seleccionados para poblar el modelo obtenido, dichos datos serán procesados y utilizados en etapas posteriores.
7. **Diseño e implementación del buscador semántico:** La etapa final del proyecto involucra el diseño e implementación de un prototipo de buscador semántico que permita realizar búsquedas sobre el modelo de integración.

1.4. Estructura de la tesis

El presente proyecto de investigación se encuentra estructurado en seis capítulos, además de las secciones de anexos y bibliografía. La estructura es la siguiente:

1. **Introducción:** En este capítulo se describe en qué consiste el proyecto de investigación, presentando su justificación, objetivos, metodología de investigación, además de la estructuración del presente documento.
2. **Fundamentos teóricos:** Este capítulo contiene conceptos teóricos fundamentales para la comprensión del presente trabajo, adicionalmente se describen el ciclo de vida y la metodología utilizadas para el desarrollo del modelo de integración de datos.
3. **Investigaciones relacionadas:** Ese capítulo presenta los modelos propuestos por cuatro trabajos centrados en describir áreas específicas del dominio universitario,



además como resultados obtenidos, se describe el aporte de cada uno de estos trabajos al desarrollo del presente proyecto.

4. **Proceso de integración de datos:** En este capítulo se describe el proceso llevado a cabo para integrar las diversas fuentes de datos de la Universidad de Cuenca, e incluye cada una de las etapas especificadas por el ciclo de vida seleccionado, donde se considera también el proceso de modelado ontológico haciendo uso de la metodología establecida.
5. **Prototipo del buscador:** Este capítulo pretende describir el prototipo del buscador desarrollado así como su funcionamiento y la presentación de resultados obtenidos al ejecutar una búsqueda específica.
6. **Conclusiones y trabajos futuros:** En este capítulo se presentan las consideraciones a destacar de los resultados obtenidos en este trabajo, además se plantean las tareas que pueden ser realizadas en un futuro para mejorar o extender el proyecto desarrollado.

Capítulo 2

Fundamentos teóricos

2.1. Evolución de la web

La web contiene una gran cantidad de información que crece día a día, por lo tanto es objeto de estudio y evolución. Desde su creación en 1989 por Tim Berners Lee, las tecnologías web han evolucionado hasta permitir a los desarrolladores crear nuevas experiencias web, es así que de una web basada en documentos y enlaces de hipertexto se ha llegado a una web que brinda soporte para archivos multimedia, inferencia de datos, etc.

A continuación se resumen las principales características de cada una de las etapas de la evolución de la web[9, 10, 11, 12].

Web 1.0: “*La Web*”: Nace como un sistema de hipertexto para compartir información en internet, con la finalidad de publicar documentos.

- Pocos productores de contenidos.
- Los usuarios actúan únicamente como lectores de los contenidos, no pueden afectar o contribuir a los sitios.
- El contenido en los sitios es estático ya que no se realiza una actualización en forma periódica.

Web 2.0: “*Web Social*”: Basada en comunidades de usuarios, marca un cambio de una web informativa creada por expertos a una web social donde cualquiera puede participar fácilmente.

- Los usuarios se convierten en contribuidores. Publican información y realizan cambios en los datos.

- No se requieren grandes conocimientos técnicos para crear un espacio de internet con contenidos que puedan ser accedidos por otros usuarios.

Web 3.0: “Web Semántica”: Basada en la representación del significado de la información [13].

- Añade metadatos semánticos a la información de la web a través de la utilización de ontologías.
- Debido a la semántica agregada se hace posible acceder a la información requerida de forma exacta y completa.
- En la web semántica se crean enlaces, por lo que una persona o máquina puede explorar la red de datos.

2.2. Web Semántica

Tomando en consideración que la gran cantidad de información disponible que existe actualmente en internet no se encuentra clasificada correctamente, los resultados obtenidos de las búsquedas realizadas por un usuario son inexactas y en algunos casos irrelevantes. Para dar solución a este inconveniente se desarrolla la web semántica, la cual, es una extensión de la web actual donde la información es dotada de un significado bien definido haciendo posible su comprensión tanto por personas como por las máquinas.

La web semántica está dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la web de más significado se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de un modelo común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante[13].

2.2.1. Arquitectura

De acuerdo a Tem Berners-Lee la arquitectura de la web semántica presenta diversas capas de abstracción tal como se muestra en la figura 2.1; a continuación se detallará cada una de ellas.

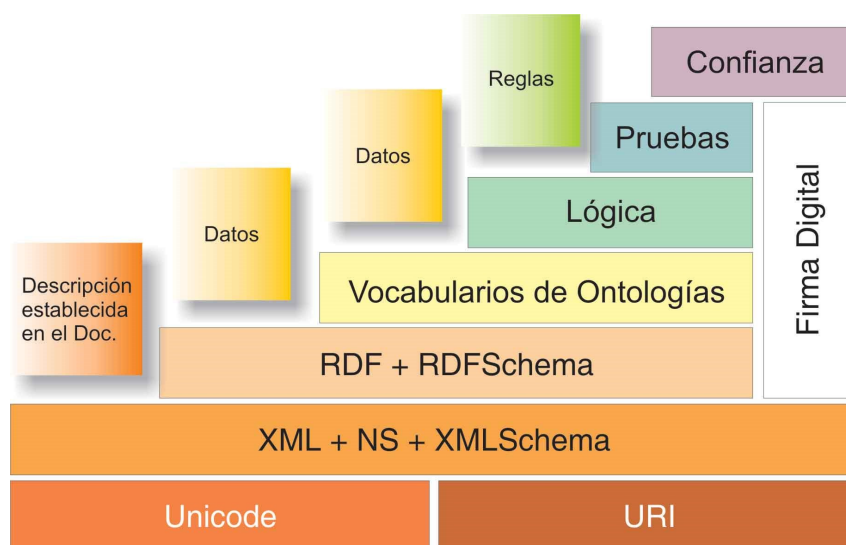


Figura 2.1: Arquitectura de la web semántica[1].

Unicode - URI

- **Unicode:** Estándar de codificación de caracteres universal utilizado para la representación de texto[14].
- **URI (*Unified Resource Identifier*):** Cadena de caracteres que permite identificar de forma única un recurso en la web[15].

Esta capa básica de abstracción hace referencia al nombrado e identificación única de cada uno de los recursos y codificación de los mismos[5].

XML + NS + XMLSchema

- **XML (*Extensible Markup Language*):** Lenguaje que permite describir datos mediante etiquetas, además define una estructura de árbol procesable por las máquinas[16].
- **NS (*Namespace*):** Permite utilizar descripciones realizadas en documentos externos.
- **XML-Schema:** Lenguaje de esquema que permite describir la estructura y restricciones de un documento XML[17].

Si bien en este nivel de abstracción se establece un lenguaje fácilmente procesable que permite la descripción de recursos (XML), este no permite realizar inferencia y no es posible definir un significado de las etiquetas definidas.

RDF + RDF *Schema*

- **RDF (*Resource Description Framework*):** Lenguaje que permite definir la sintaxis y modelos de datos para la representación semántica de los datos[18].
- **RDF-*Schema*:** Se construye sobre RDF y sirve para definir propiedades y tipos de recursos, además permite definir vocabularios con términos y relaciones entre los mismos[19].

En esta capa RDF se encarga de definir los formalismos necesarios para anotar metadatos en sintaxis XML y por otra parte RDF-*Schema* permite proporcionar cierta semántica a los recursos y propiedades haciendo posible que las máquinas puedan realizar ciertas tareas[20].

Vocabularios de ontologías: Consiste en una representación formal de los conceptos existentes en un dominio específico y las relaciones entre ellos.

Lógica: Interpretación común de las operaciones realizadas a nivel ontológico.

Pruebas: Asegurar la corrección de los razonamientos y deducciones.

Confianza: Confiabilidad de la información en la web semántica.

2.3. Ontologías

Para construir la web semántica es necesario representar el conocimiento de forma que sea legible por los ordenadores, de manera que sea consensuado y reutilizable. Las ontologías proporcionan la vía para lograr este objetivo y hacen posible que los sistemas de recuperación de información trabajen con conceptos en vez de palabras clave, lo cual a su vez permite obtener información de forma precisa y poder realizar deducciones[21].

Existen diversas definiciones de ontología en el contexto de la web semántica, en este trabajo se considera la definición presentada por la W3C:

”Una ontología define los términos utilizados para describir y representar un área de conocimiento. Las ontologías son utilizadas por personas, bases de datos y aplicaciones que necesitan compartir información de un dominio (un dominio es sólo un área temática específica o área de conocimiento, como medicina, fabricación de herramientas, bienes raíces, gestión financiera, etc.). Las ontologías incluyen definiciones de conceptos básicos en el dominio y las relaciones entre ellas”[22].

En resumen, una ontología permite representar el conocimiento de un dominio seleccionado, definiendo un vocabulario formal y estándar que describe los conceptos que contiene y las relaciones existentes entre dichos conceptos. Además *“una ontología recoge reglas lógicas y restricciones para hacer “comprender” a las máquinas los conceptos que manejan dentro de un determinado campo”*[23]. Por ejemplo, una ontología que describe el dominio de educación primaria establece que todos los profesores son personas, pero no todas las personas son profesores (ya que una persona puede ser un estudiante o un empleado).

2.3.1. Elementos básicos de una ontología

Para facilitar la comprensión de los elementos básicos que conforman una ontología en la figura 2.2 se ilustra un ejemplo básico de una ontología para el dominio educativo. A continuación se describirán cada uno de los elementos.

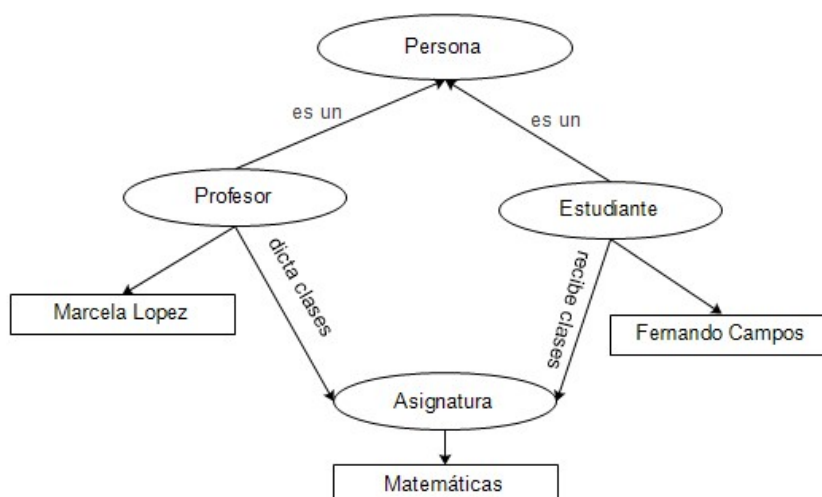


Figura 2.2: Ejemplo ilustrativo de los elementos de una ontología

- **Conceptos:** Representan las ideas básicas que se intentan formalizar[24]. En el ejemplo planteado se definen los conceptos “Persona” y “Asignatura”. Los conceptos “Profesor” y “Estudiante” se han creado como especializaciones (subclases) de “Persona”.
- **Relaciones:** Representan la interacción y enlace entre conceptos del dominio (jerarquías)[25]. En este caso las relaciones “es un” genera una relación de subclase y permite definir entidades complejas a partir de entidades más simples de manera incremental. Adicionalmente, se han definido las propiedades “dicta clases”

y “recibe clases” que permiten relacionar los conceptos “Profesor” y “Estudiante” respectivamente con “Asignatura”.

- **Instancias:** Se utilizan para representar objetos determinados de un concepto. En el ejemplo considerado se han definido en rectángulos las instancias de cada uno de los conceptos.

2.4. Linked Open Data

Con el objetivo de hacer más comprensible el concepto de Datos Abiertos Enlazados (*Linked Open Data* - LOD) se definirá previamente los conceptos de Datos Enlazados (*Linked Data*) y Datos Abiertos (*Open Data*) de forma independiente.

La web semántica no involucra únicamente la publicación de datos en la web, sino que éstos puedan ser vinculados a otros, permitiendo la exploración de datos tanto a las personas como a las máquinas. Para lograr dicha vinculación se establecen los datos enlazados que permiten vincular los distintos datos distribuidos en la web, de forma que puedan ser referenciados de la misma forma que lo hacen los enlaces de las páginas web.

Los Datos Enlazados, como parte de la web semántica, se basan en la aplicación de cuatro principios básicos y necesarios [26]:

1. Utilizar URIS para identificar recursos de forma estándar y unívoca.
2. Publicar los recursos haciendo uso de una URI basada en HTTP para asegurar que cualquier recurso pueda ser buscado y accedido en la web.
3. Proporcionar información útil descrita en RDF una vez que un recurso ha sido accedido mediante una URI basada en HTTP.
4. Incluir enlaces a otros URIs relacionadas con los datos contenidos en el recurso, de forma que no se encuentren aislados y así se pueda compartir información con otras fuentes externas.

Por otra parte, Datos Abiertos es una filosofía y práctica que tiene por objetivo poner a disposición de toda la sociedad determinados datos e información pertenecientes a administraciones públicas, en formatos fáciles de manipular, de manera que puedan ser analizados y reutilizados por cualquier persona o empresa. Poner los datos a disposición de la sociedad hace que cualquier persona, empresa u organización pueda reutilizarlos y con ello generar nuevos datos, conocimientos, mejorar procesos; en general agregar un valor añadido a los datos existentes[5].

En base a las definiciones anteriores se puede establecer que LOD comprende dos conceptos clave, el primero hace referencia a que los datos en la web deben encontrarse enlazados de tal forma que puedan ser accesibles tanto para personas como para computadoras. El segundo concepto involucra que para hacer posible la vinculación y reutilización de los datos éstos deben ser abiertos, es decir encontrarse libres de restricciones y derechos de *copyright*.

2.5. Vocabularios ontológicos

Con la finalidad de conseguir que los recursos existentes en la web semántica sean reutilizables, es indispensable que el significado de un recurso esté claramente definido, para lograr este objetivo, es necesario utilizar vocabularios de referencia o vocabularios de uso común, los cuales generalmente poseen una semántica bien definida.

Al igual que en la comunicación cotidiana un grupo particular de personas utiliza vocabularios específicos en sus conversaciones (por ejemplo, médicos que hablan sobre un determinado diagnóstico de un paciente haciendo uso de términos de un vocabulario relacionado con su área), en la web semántica para cada dominio específico debe realizarse una búsqueda de vocabularios existentes que puedan ser aplicados. En el caso de que ninguno de los vocabularios satisfaga las necesidades de expresión requeridas, puede crearse una nueva ontología, siempre teniendo en cuenta la reutilización de elementos existentes en otras ontologías[27].

Actualmente existe una gran cantidad de vocabularios destinados a describir diversos dominios, por ejemplo:

- **FOAF (Friend Of A Friend)**¹: Permite describir personas, sus actividades y relaciones con otras personas y objetos
- **VCARD**²: Permite describir una libreta de contactos con información básica de cada una de las personas u organizaciones registradas.
- **Time ontology**³: Proporciona un vocabulario que permite expresar tiempos e intervalos junto con información sobre duración y posición temporal, incluyendo fecha y hora.

¹<http://xmlns.com/foaf/0.1/>

²<http://www.w3.org/2006/vcard/ns#>

³<http://www.w3.org/2006/time#>

- **SCHEMA**⁴: Proporciona una colección de vocabularios compartidos que permiten representar eventos, organizaciones, lugares y productos.
- **DBPEDIA**⁵: Contiene una ontología correspondiente a los datos extraídos de las nuevas versiones de Wikipedia.

Cada vocabulario existente está descrito por un documento referenciado por una URI. Para referenciar a cada una de las clases y propiedades dicha URI está construida uniendo la URI del vocabulario con el nombre de la respectiva clase o propiedad. Tomando como ejemplo el vocabulario FOAF:

- URI para referenciar el vocabulario: “<http://xmlns.com/foaf/0.1/>”
- URI de la propiedad edad: “<http://xmlns.com/foaf/0.1/age>”

2.6. Tecnologías semánticas

Para lograr que el conocimiento a representar pueda ser comprendido y procesado por máquinas es necesario expresarlo de forma adecuada, es por esto que la W3C lleva años trabajando en diversos lenguajes, en su mayoría basados en la sintaxis XML, que están definidos para ser capaces de representar ontologías y vocabularios mediante tripletas y con algunas construcciones de mayor complejidad[28].

2.6.1. RDF (Resource Description Framework)

RDF fue definido por la W3C con la intención de crear un lenguaje de descripción de metadatos[28]. Provee un método general y flexible para descomponer el conocimiento en partes pequeñas llamadas tripletas, debido a que se componen de tres elementos: sujeto, predicado y objeto[29].

Una sentencia particular puede ser representada haciendo uso de una tripleta (sujeto, predicado, objeto) o mediante grafos dirigidos. Así en RDF es posible representar declaraciones simples como un grafo de nodos y arcos que representan los recursos, sus propiedades y valores. En la figura 2.3 se ilustra la representación en forma de grafo de la sentencia “*El profesor se llama Juan Perez*”.

⁴<https://schema.org/docs/schemas.html>

⁵<http://dbpedia.org/ontology/>

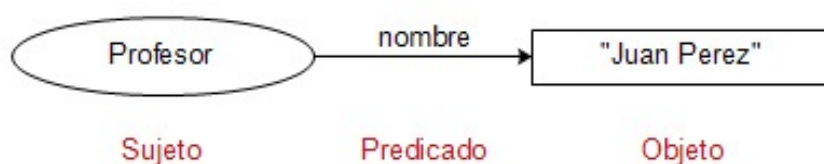


Figura 2.3: Estructura de una tripleta RDF

- **Sujeto:** Recurso, objeto o cosa a la que se hace referencia en una afirmación.
- **Predicado:** Describe la relación existente entre dos recursos.
- **Objeto:** Valor del predicado para el sujeto.

Debido a que RDF se basa en los estándares de URI y Unicode, las tripletas pueden ser representadas haciendo uso de URIs que identifican a los recursos utilizados en la afirmación. En la tabla 2.1 se representa mediante URIs el ejemplo descrito en la figura 2.3, para esto se ha hecho uso de los vocabularios FOAF y TEACH⁶. En este caso, debido a que el objeto de la tripleta es un valor literal, no requiere una URI para su representación, pero existen casos en los que el objeto puede ser referenciado por una URI, esto indica que el sujeto de una declaración puede ser el objeto de otra, por lo tanto su representación formará un grafo dirigido.

Tabla 2.1: Ejemplo de representación de una tripleta mediante URIs

Sujeto	Predicado	Objeto
Profesor	nombre	Juan Perez
<i>http://linkedscience.org/teach/ns/#teacher</i>	<i>http://xmlns.com/foaf/0.1/name</i>	<i>"Juan Perez"</i>

2.6.2. RDF – Schema (RDFS)

RDF – Schema proporciona el marco para describir clases y propiedades específicas de una aplicación. Las clases definidas en RDFS son similares a las clases creadas en lenguajes de programación orientados a objetos. Esto permite que los recursos se definan como instancias de clases y subclases de clases[19].

RDFS está construido sobre RDF y permite básicamente:

- Modelar objetos con una semántica explícita.
- Especificar restricciones de tipos de datos.

⁶TEACH (Teaching Core Vocabulary): Provee terminos que permiten relacionar profesores con cosas de un curso.

- Definir el significado, características y relaciones entre propiedades.
- Definir dominio y rango de propiedades.
- Establecer jerarquías de clases.

En la tabla 2.2 se presentan algunas de las clases y propiedades contenidas en RDFS.

Tabla 2.2: Clases y propiedades contenidas en RDF-Schema[5].

Tipo	Nombre	Descripción
Clase	rdfs:Class	Permite definir una clase
	rdf:Property	Permite definir una propiedad RDF.
	rdfs:Literal	Clase de valores literales.
Propiedad	rdfs:Domain	Permite indicar que cualquier recurso que tiene una propiedad determinada es instancia de una o más clases.
	rdfs:range	Permite indicar que los valores de una propiedad son instancias de una o más clases.
	rdfs:subClassOf	Permite indicar que todas las instancias de una clase son instancias de otra.
	rdfs:subPropertyOf	Indica que el sujeto es una subpropiedad de una propiedad.

2.6.3. SPARQL (Simple Protocol and RDF Query Language)

SPARQL es un lenguaje de consulta utilizado en el ámbito de la web semántica. Define la sintaxis y semánticas necesarias para expresar una consulta sobre un grafo RDF y los diferentes tipos de resultados que pueden ser obtenidos (conjuntos de tripletas RDF, grafos RDF, URIs de recursos o valores)[30].

Al ser un lenguaje de consulta, SPARQL tiene una sintaxis establecida. Los elementos de dicha sintaxis son descritos a continuación[5]:

- **PREFIX:** Asocia una URI a una etiqueta que será utilizada posteriormente en la consulta. La sentencia utilizada para esto es `PREFIX <prefijo>:<URI>`.
- **SELECT:** Es similar al uso especificado por el estándar SQL. Permite definir las variables que deben ser devueltas como resultados, para esto los nombres de cada una de las variables deben iniciar con '?' y estar separadas por ','.
- **FROM:** Identifica los datos sobre los cuales será ejecutada la consulta.

- **WHERE:** Define un patrón de consulta, con una o más tripletas encerradas entre llaves.

En la figura 2.4 se presenta una consulta simple definida en lenguaje SPARQL que obtiene como resultado los nombres y direcciones de correo de todas las personas contenidas en un grafo RDF.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
{ ?x foaf:name ?name .
  ?x foaf:mbox ?mbox }
```

Figura 2.4: Ejemplo de consulta en SPARQL

2.6.4. OWL (Web Ontology Language)

OWL es un lenguaje de etiquetado semántico utilizado para publicar y compartir ontologías en la web. Ha sido desarrollado como una extensión de RDF y se deriva del lenguaje DAML+OIL ⁷. Posee mayores funcionalidades para expresar el significado y semántica que XML, RDF y RDFS, para esto añade más vocabulario que permite describir propiedades, clases, cardinalidad, igualdad, características de propiedades, clases enumeradas, etc. Adicionalmente OWL ofrece la posibilidad de que el contenido representado sea procesable por máquinas[31].

OWL ofrece tres sub-lenguajes de expresión incremental diseñados para ser utilizados por comunidades específicas de desarrolladores y usuarios según el nivel de expresividad que se requiera[32].

- **OWL Lite:** Dirigido a usuarios que primordialmente necesitan una clasificación jerárquica y restricciones simples.
- **OWL DL:** Da soporte a usuarios que requieren la máxima expresividad mientras se garantiza que todas las conclusiones son computables y que todas las operaciones realizadas terminarán en un tiempo finito.
- **OWL Full:** Da soporte a usuarios que requieren la máxima expresividad y libertad sintáctica de RDF pero sin garantías computacionales. Se debe tener en consideración que es poco probable que algún software racional sea capaz de soportar completamente el razonamiento para cada una de las características de OWL Full.

⁷DAM+OIL es un lenguaje producto de la unificación de los lenguajes DAML (DARPA's Agent Markup Language) y OIL (Ontology Inference Layer).

Con la finalidad de agregar nuevas características como mayor expresividad sintáctica, extensiones sobre los tipos de datos y capacidad de anotación se define OWL 2, que es un lenguaje enfocado específicamente en definir ontologías para la web. Al igual que su predecesor, OWL 2 define los siguientes sublenguajes[33]:

- **OWL 2 EL:** Útil en aplicaciones que hacen uso de ontologías que contienen un gran número de clases y/o propiedades, y garantiza que todas las tareas de razonamiento comunes sean ejecutadas en un tiempo polinomial con respecto al tamaño de la ontología.
- **OWL 2 QL:** Dirigido a aplicaciones que utilizan grandes volúmenes de datos de instancias, y donde la respuesta a consultas es la tarea de razonamiento de mayor importancia. Permite implementar consultas conjuntivas⁸ mediante sistemas de bases de datos relacionales.
- **OWL 2 RL:** Es un sublenguaje que permite ejecutar razonamientos sobre la ontogías en tiempo polinomial utilizando técnicas de base de datos. Aplicable especialmente a aplicaciones que requieren un razonamiento escalable sin sacrificar demasiada potencia expresiva.

2.7. Ciclos de vida para datos enlazados

Existen diversos ciclos de vida aplicables para la publicación de proyectos LOD, en esta sección se presentan tres de ellas, las cuales han sido tomadas de la guía de buenas prácticas para la publicación de datos enlazados⁹.

2.7.1. Ciclo de vida de Hyland

En este ciclo de vida se definen siete pasos para modelar, crear, publicar y anunciar datos enlazados, dicho ciclo de vida consiste en las actividades descritas en la figura 2.5.

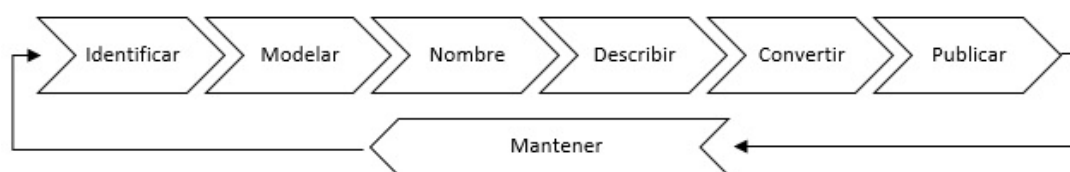


Figura 2.5: Actividades propuestas en el ciclo de vida de Hyland[2]

⁸Consulta conjuntiva: Forma de consultas que utilizan el operador de conjunción lógica.

⁹W3C, Best Practices for Publishing Linked Data. <https://www.w3.org/TR/ld-bp/>

A continuación se presenta un resumen de los pasos contenidos en este ciclo de vida, tal y como se describe en *Linked Data Cookbook*[34]:

1. **Identificar:** Involucra actividades relacionadas con la revisión de los recursos disponibles y la búsqueda de objetos de interés real como personas, lugares y cosas.
2. **Modelar:** Consiste en realizar una descripción gráfica de los objetos y las relaciones que existen entre ellos. En esta etapa se recomienda investigar trabajos donde se hayan descrito datos relacionados o similares con la finalidad de reutilizar vocabularios comunes.
3. **Nombre:** Haciendo uso de URIs se deben definir nombres para cada uno de los objetos.
4. **Describir:** Crear un esquema que describa adecuadamente el dominio considerado, reutilizando vocabularios siempre que sea posible.
5. **Convertir:** Una vez que se tenga un esquema bien definido se deberá convertir los datos de origen en una representación de datos enlazados.
6. **Publicar:** Copiar el conjunto de datos en un servidor web accesible públicamente.
7. **Mantener:** Los editores deben ser responsables de mantener los datos actualizados, asegurar su exactitud y reparar posibles problemas reportados.

2.7.2. Ciclo de vida de Hausenblas

En este ciclo de vida se considera que los enfoques existentes de gestión de datos asumen el control sobre los datos, su esquema y generación, esto no ocurre en la web debido a que se trata de un entorno abierto y descentralizado. Las seis actividades propuestas en este modelo se ilustran en la figura 2.6.



Figura 2.6: Actividades establecidas en el ciclo de vida de Hausenblas [2]

A continuación se presenta una descripción resumida de cada una de las actividades contenidas en este ciclo de vida[6]:

- **Conocimiento de datos:** Esta actividad consiste en definir el dominio y los conjuntos de datos existentes.

- **Modelamiento:** Refiere a las tareas de modelar los datos definiendo una ontología apropiada.
- **Publicación:** Esta actividad se centra en la transformación de los datos de origen a RDF y su posterior publicación en la web.
- **Descubrimiento:** Involucra el desarrollo de un modelo para la descripción de metadatos sobre los conjuntos de datos RDF.
- **Integración:** En esta actividad se crean enlaces entre los diversos conjuntos de datos establecidos.
- **Casos de uso:** Consiste en definir escenarios de casos de uso o desarrollo de aplicaciones y servicios específicos que aprovechan el conjunto de datos enlazados.

2.7.3. Ciclo de vida de Villazón - Terrazas

Villazón - Terrazas proponen un ciclo de vida para generar, publicar y exportar datos gubernamentales enlazados. Las actividades principales definidas en este modelo se presentan en la figura 2.7, pero cada una de ellas se descomponen en una o varias tareas y se proporciona algunas técnicas y herramientas para llevar a cabo[35].

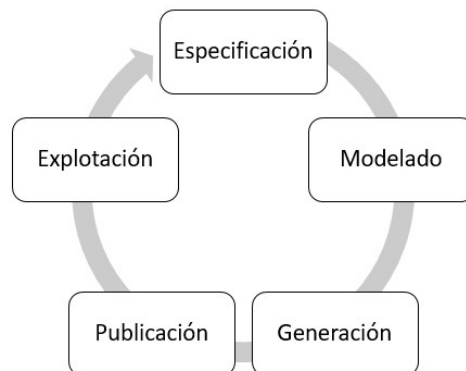


Figura 2.7: Actividades definidas en el ciclo de vida de Villazón – Terrazas[2]

1. **Especificación:** Esta primera actividad consiste en elaborar una especificación detallada de los requisitos, esto permitirá establecer una base para la estimación de costos y cronogramas además de proveer un método para la validación y verificación.

Las tareas identificadas para esta actividad son:

- a) **Identificación y análisis de fuentes de datos gubernamentales:** En esta tarea se identifican, analizan y seleccionan los datos que se desean publicar.

Es necesario distinguir si se requieren publicar datos que aún no han sido publicados o reutilizar y aprovechar datos ya publicados por agencias gubernamentales.

- b) **Diseño de URIs:** Para lograr la visión de la web como una base de datos global se requiere identificar los recursos mediante URIs, las cuales deben ser diseñadas teniendo en cuenta que deben ser simples, estables y manejables, ya que una URI es un identificador no un nombre para un recurso web.
 - c) **Definición de licencia:** Al trabajar dentro de un contexto gubernamental es importante definir el tipo de licencia que tendrán los datos que se están publicando.
2. **Modelado:** Una vez realizada la especificación se requiere determinar la ontología que será utilizada para modelar el dominio de las fuentes de datos identificadas. En esta etapa se debe considerar un enfoque basado en la reutilización ya que este acelera el desarrollo de la ontología y permite ahorrar tiempo, esfuerzo y recursos. Esta actividad involucra las siguientes tareas:
 - a) Buscar vocabularios que pueden ser reutilizados, para esto existen diversos repositorios como Swoogle¹⁰, Watson¹¹ y LOV¹².
 - b) Si no se ha encontrado un vocabulario que pueda ser reutilizado se procederá a crear uno nuevo, para esto se tratará de reutilizar tanto como sea posible los recursos existentes.
 - c) Finalmente, si no existen vocabularios o recursos para construir la ontología requerida, se tendrá que crear una ontología desde cero.
3. **Generación:** Las fuentes de datos seleccionadas en la especificación son transformadas en RDF de acuerdo al vocabulario establecido en el modelado. Esta actividad incluye la siguientes tareas:
 - a) **Transformación:** El contenido de las fuentes de datos es transformado a instancias RDF. Algunos requerimientos de transformación que deben ser considerados son:
 - **Conversión completa:** Implica que las consultas que pueden ser realizadas en las fuentes deberían poder ser realizadas en la versión RDF.
 - Las instancias RDF generadas deben reflejar la estructura ontológica de la forma más aproximada posible.

¹⁰<http://swoogle.umbc.edu/2006/>

¹¹<http://watson.kmi.open.ac.uk/WatsonWUI/>

¹²LOV (Linked Open Vocabularies): <http://lov.okfn.org/dataset/lov/vocabs>

- b) **Limpieza de datos:** Consiste en identificar, encontrar y corregir errores comunes y vínculos rotos en los datos que impidan a las aplicaciones explotar la información de forma eficiente.
 - c) **Enlazado:** Esta tarea se centra en incluir enlaces entre los conjuntos de datos gubernamentales y los conjuntos de datos externos. Los pasos a llevar a cabo para esto son:
 - Identificar los datos que pueden ser apropiados para la vinculación.
 - Descubrir las relaciones existentes entre el conjunto de datos gubernamentales y los datos identificados.
 - Validar las relaciones descubiertas.
4. **Publicación:** En actividad refiere a la publicación de los datos RDF y consta de las siguientes tareas:
- a) Publicación del conjunto de datos.
 - b) Publicación de los metadatos.
 - c) Permitir el descubrimiento y la sincronización efectivos del conjunto de datos.
5. **Explotación:** El objetivo final de la publicación de datos es fomentar el uso y la reutilización de esta información. Por lo tanto es necesario desarrollar aplicaciones que exploten estos datos y proporcionen interfaces gráficas para los usuarios.

2.7.4. Selección de ciclo vida a utilizar

En base a lo descrito se puede establecer que los tres ciclos de vida considerados comparten actividades comunes, es decir, contienen actividades similares requeridas en el proceso de publicación; pero establecen límites diferentes, tal como se muestra en la tabla 2.3, por lo tanto no se puede considerar que un determinado ciclo de vida es mejor que otro, sino más bien, cada una tiene una manera diferente de establecer las actividades del proceso de gestión de datos enlazados.

En el caso particular del presente trabajo, se ha decidido utilizar el ciclo de vida propuesto por Villazón – Terrazas, por lo siguiente:

- El ciclo de vida propuesto por Villazón - Terrazas ha sido ampliamente utilizado en trabajos previos, por lo cual existe documentación que puede ser consultada.
- Como resultado de proyectos de investigación previos en la Universidad de Cuenca se han desarrollado herramientas que permiten automatizar varias de las tareas incluidas en las diferentes actividades del ciclo de vida, lo cual permite ahorrar un tiempo considerable en el desarrollo de proyectos LOD.

Tabla 2.3: Alcance de actividades definidas en los diferentes ciclos de vida[6]

Objetivo	Ciclo de vida		
	Hyland	Hausenblas	Villazón-Terrazas
Conocimiento del dominio y los datos	Identificación	Conocimiento de datos	Especificación
Modelamiento de los datos	<ul style="list-style-type: none">■ Modelar■ Nombrar	Modelamiento	Modelado
Transformación a RDF	<ul style="list-style-type: none">■ Describir■ Convertir	<ul style="list-style-type: none">■ Publicación■ Descubrimiento■ Integración	Generación
Publicación de los datos en la web	Publicar		Publicación
Casos de uso, aplicaciones y servicios		Casos de uso	Explotación

2.8. Metodología NeOn para el desarrollo ontológico

La construcción de ontologías puede verse como un proceso similar a la construcción de software, es por ello que requiere una metodología que guíe este proceso. En esta sección se describe la metodología NeOn, la cual se utiliza en el desarrollo de la red de ontologías UCuenca-Ontology.

2.8.1. El marco de trabajo de la metodología NeOn

En la web semántica las ontologías no se encuentran aisladas unas de otras, sino más bien se relacionan entre si formando una red de recursos semánticos interconectados conocida como red de ontologías o red ontológica. Es por esto que la metodología NeOn utiliza un enfoque basado en escenarios para el desarrollo ontológico y provee un conjunto integral de métodos y directrices para llevar a cabo las actividades requeridas para el desarrollo de ontologías en red[4]. Además, cada uno de los procesos o actividades son definidos de forma precisa, indicando su propósito, entradas, salidas, actores involucrados, cuándo su ejecución es más apropiada, además los métodos, herramientas y técnicas propuestos para ser utilizados.

"Metodologías actuales como METHONTOLOGY y DILIGENT plantean directrices para la construcción de una sola ontología que va desde la especificación de requerimientos hasta la implementación de la ontología, en cambio la metodología NeOn no establece un flujo de trabajo rígido sino que sugiere caminos y actividades para diferentes escenarios que cubren los situaciones más comunes que suelen ocurrir"[4].

2.8.2. Escenarios para construir redes ontológicas

La metodología NeOn plantea nueve escenarios flexibles para la construcción de ontologías y redes ontológicas, poniendo especial énfasis en la reutilización y re-ingeniería de recursos de conocimiento (ontológicos y no ontológicos) que hayan sido identificados.[3]

De acuerdo a lo planteado en *Ontology Engineering in a Networked World*[4] y en *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse*[3], a continuación se describe los casos en los cuales se debe aplicar un escenario, sus actividades principales y el resultado obtenido en cada caso.

2.8.2.1. Escenario 1: De la especificación a la implementación

Aplicar

- En el desarrollo de ontologías desde cero.
- Debe ser combinado con el resto de escenarios ya que contiene las actividades principales que deben ser realizadas en cualquier desarrollo ontológico.

Actividades principales

1. **Especificación de requerimientos ontológicos**
2. **Planificación:** Se establece el ciclo de vida y recursos necesarios.
3. **Conceptualización:** El conocimiento es organizado y estructurado en modelos.
4. **Formalización:** El modelo conceptual es transformado en un modelo semi-computable.
5. **Implementación:** El modelo obtenido es implementado en un lenguaje ontológico.

Salida

1. Red ontológica que representa el dominio implementado en un lenguaje ontológico.
2. Varios documentos como: Documento de especificación de requerimientos, documento de descripción de la ontología, documento de evaluación de la ontología.

2.8.2.2. Escenario 2: Reutilización y re-ingeniería de recursos no ontológicos

Aplicar

- Cuando los desarrolladores disponen de recursos no ontológicos que pueden ser re-utilizados.
- Se debe considerar que al reutilizar recursos no ontológicos involucra necesariamente su re-ingeniería en ontologías.

Actividades principales

1. **Reutilización de recursos no ontológicos**
 - a) Buscar y analizar los recursos no ontológicos (bases de datos, archivos, etc.)
 - b) Evaluar el conjunto de recursos candidato
 - c) Seleccionar los recursos apropiados.

2. Re-ingeniería de recurso no ontológicos

- a) **Ingeniería inversa:** Analizar los recursos para identificar los componentes y crear representaciones a diferentes niveles de abstracción.
- b) **Transformación:** Generar un modelo conceptual de recurso no ontológico.
- c) **Ingeniería avanzada:** Implementar la ontología en base al modelo identificado en la actividad anterior.

Salida

1. Salidas del escenario 1.
2. Recursos no ontológicos transformados por medio de la actividad de re-ingeniería.

2.8.2.3. Escenario 3: Reutilizar recursos ontológicos

Aplicar

- Cuando los desarrolladores tienen a su disposición recursos ontológicos útiles para el problema y que pueden ser reutilizados.

Actividades principales

1. **Búsqueda:** Buscar recursos ontológicos candidatos que satisfagan los requerimientos.
2. **Evaluación:** Verificar que el contenido y granularidad del recurso obtenido sea el adecuado.
3. **Comparación:** Los recursos obtenidos deben ser comparados en base a criterios identificados por los desarrolladores.

Salida

- Salidas del escenario 1.

2.8.2.4. Escenario 4: Reutilización y re-ingeniería de recursos ontológicos

Aplicar

- Utilizado cuando los desarrolladores disponen de recursos ontológicos que pueden ser reutilizados, pero requieren ser modificados para servir al problema o propósito.

Actividades principales

1. Actividades del escenario 3.
2. **Re-ingeniería de recursos ontológicos:** El proceso de re-ingeniería puede llevarse a cabo en distintos niveles de abstracción (especificación, conceptualización, formalización e implementación). La forma concreta en que se realice dicho proceso dependerá de las características que necesitan ser modificadas en el recurso ontológico.

Salida

1. Salidas del escenario 1
2. Recursos ontológicos nuevos seleccionados para ser reutilizados y generados mediante el proceso de re-ingeniería. Algunos de estos nuevos recursos pueden ser considerados como nuevas versiones de los recursos ontológicos luego del proceso de re-ingeniería.

2.8.2.5. Escenario 5: Reutilización y fusión de recursos ontológicos**Aplicar**

1. Considera los casos en los que se disponen de varios recursos en el mismo dominio que pueden ser reutilizados y se desea crear uno nuevo partiendo de dos o más.

Actividades principales

1. Actividades del escenario 3.
2. **Actividades de fusión de ontologías**
 - a) **Alineación:** Obtener un conjunto de alineamientos entre los recursos seleccionados.
 - b) **Fusión:** Fusionar los recursos utilizando los alineamientos obtenidos en la actividad anterior, para obtener un nuevo recurso.

Salida

1. Un conjunto de alineamientos entre los recursos ontológicos seleccionados.
2. Un conjunto de recursos ontológicos nuevos a ser integrados en la red ontológica.

2.8.2.6. Escenario 6: Reutilización, Fusión y re-ingeniería de recursos ontológicos

Aplicar

- Cuando se fusionan varios recursos ontológicos para formar la red ontológica, pero esto no es útil tal como está y requiere ser modificado.

Actividades principales

1. Actividades del escenario 3.
2. Actividades de fusión de recursos ontológicos del escenario 5.
3. Actividades de re-ingeniería de recursos ontológicos del escenario 4.

Salida

1. salidas del escenario 1
2. Recurso ontológico fusionado.
3. Recurso ontológico fusionado y con re-ingeniería.
4. Alineamientos entre los recursos ontológicos seleccionados.

2.8.2.7. Escenario 7: Reutilización de patrones de diseño (ODPs)

Aplicar

- Involucra aquellos casos en los que es posible aplicar las mejores prácticas para el desarrollo de la red ontológica.

Actividades principales

- Durante el desarrollo de una ontología es común que surjan dudas sobre la forma en que un determinado conocimiento debe ser modelado. En este caso se pueden acceder a librerías que ofrecen soluciones de modelado.

Salida

- Un conjunto de patrones de diseño ontológico integrados en la red ontológica desarrollada.

2.8.2.8. Escenario 8: Re-estructuración de recursos ontológicos

Aplicar

- En los casos en los que el conocimiento contenido en un modelo conceptual de la red ontológica debe ser corregido y reorganizado para obtener una red que cumpla los requerimientos de la ontología.

Actividades principales

- **Re-estructuración ontológica:** Puede realizarse de forma independiente o como parte de un proceso de re-ingeniería, ejecutando una o varias de las siguientes actividades:
 1. **Modularización:** Crear diferentes módulos facilita la reutilización del conocimiento incluido en la red.
 2. **Poda:** Quitar ramas de taxonomías incluidas en la red que se consideren innecesarias.
 3. **Enriquecimiento:** Puede realizarse de dos maneras:
 - a) **Extensión:** Incluir nuevos conceptos y relaciones.
 - b) **Especialización:** Agregar mayor granularidad a una rama incluyendo conceptos y relaciones más especializados.

Salida

- Red ontológica que representa el dominio que se pretende modelar.

2.8.2.9. Escenario 9: Localización de recursos ontológicos

Aplicar

- Cuando la red ontológica a ser desarrollada debe ser escrita en diferentes idiomas naturales.

Actividades principales

1. Seleccionar recursos lingüísticos apropiados.
2. Seleccionar las etiquetas ontológicas a ser localizadas.
3. Obtener la traducción de las etiquetas ontológicas.
4. Evaluar la traducción de las etiquetas.
5. Actualizar la ontología.

Salida

- Modelo conceptual de la red ontológica en un lenguaje natural diferente que representa el dominio.

2.8.3. Actividades de la ingeniería ontológica

La ingeniería ontológica hace referencia al conjunto de actividades que deben ser realizadas durante el proceso de construcción de una ontología, así como las metodologías, herramientas y lenguajes necesarios para su construcción. En esta sección se describen las actividades de especificación, planificación y reutilización de ontologías de dominio que serán aplicadas posteriormente.

2.8.3.1. Especificación de requerimientos ontológicos

La especificación de requerimientos es un proceso clave que debe ser realizado al iniciar el proyecto de desarrollo ontológico, este tiene la finalidad de identificar a los usuarios finales y los requerimientos que deben ser cumplidos por la ontología [36]. La ficha presentada en la figura 2.8 describe la información de esta actividad de forma práctica y sencilla.

La metodología NeOn proporciona guías descriptivas para el desarrollo de la especificación de requerimientos ontológicos[3]. Dichas guías pueden resumirse como sigue:

Tarea 1. Identificación del propósito, alcance y lenguaje de implementación de la ontología: Consiste en identificar el objetivo principal o finalidad de la ontología, su cobertura, granularidad y el grado de formalidad que será utilizado para codificar la ontología.

Tarea 2. Identificación de los usuarios finales previstos: En base a un conjunto de necesidades ontológicas y entrevistas con expertos del dominio y usuarios se establecerá un listado de los posibles usuarios finales de la ontología.

Tarea 3. Identificación de usos previstos: El objetivo de esta actividad es obtener un listado de usos previstos descritos en forma de escenarios, estos permitirán tener una idea general de los requerimientos de la aplicación en términos del conocimiento a ser representado.

Tarea 4. Identificación de requerimientos: Se plantean los requisitos o necesidades que deben ser cumplidos por la ontología, los cuales pueden ser divididos en los siguientes dos tipos:

Especificación de requisitos ontológicos	
Definición La actividad de especificación de requisitos consiste en definir una colección de requisitos que la ontología debe cumplir, por ejemplo, establecer por qué la ontología se tiene que construir, cuáles van a ser sus usos y sus usuarios previstos. La recopilación de estos requisitos se puede llevar a cabo mediante un proceso de consenso.	
Objetivo La actividad de especificación de requisitos establece por qué se debe desarrollar la ontología, cuáles son sus usos y usuarios previstos y un conjunto de requisitos que la ontología debe cumplir.	
Entrada Un conjunto de necesidades ontológicas.	Salida El documento de especificación de requisitos de la ontología (DERO u ORSD).
Quien Desarrolladores de software y de ontologías, quienes forman el equipo de desarrollo ontológico (EDO u ODT ¹⁷), en colaboración con los usuarios y los expertos del dominio.	
Cuando Esta actividad debe ser llevada a cabo en paralelo con la actividad de adquisición del conocimiento.	

Figura 2.8: Ficha descriptiva de la actividad de especificación de requerimientos ontológicos [3]

1. **Requerimientos ontológicos no funcionales:** Hacen referencia a las características, cualidades o aspectos generales no relacionados con el contenido que debe ser representado por la ontología, por ejemplo: convenciones de nombrado, lenguaje específico, uso de estándares.
2. **Requerimientos ontológicos funcionales:** Especifican el conocimiento que debe ser representado en la ontología. Pueden ser escritos en lenguaje natural en forma de preguntas de competencia (CQs) con sus respectivas respuestas.

Tarea 5. Agrupación de requisitos funcionales: Consiste en agrupar las CQs con el objetivo de identificar las partes esenciales a ser cubiertas por la ontología.

Tarea 6. Validación del conjunto de requerimientos: El objetivo de esta tarea es identificar posibles conflictos o contradicciones entre CQs y si es necesario incluir nuevas CQs. Para la validación de CQs se proponen los siguientes criterios: correctitud, completitud, consistencia, verificabilidad, comprensibilidad, no ambigüedad, realismo, modificabilidad y trazabilidad.

Tarea 7. Priorización del conjunto de requisitos: Añadir diferentes niveles de prioridad a los grupos de CQs y dentro de cada grupo a las CQs, esto permite planificar de mejor manera el desarrollo de la ontología.

Tarea 8. Extracción de terminología y su frecuencia: El objetivo de esta tarea es extraer un pre-glosario de términos que serán utilizados en la actividad de conceptualización a partir de las CQs y sus respuestas.

En la figura 2.9 se presenta la plantilla propuesta por la metodología NeOn para el documento de especificación de requerimientos ontológicos (ORSD), donde se incluyen las salidas obtenidas de cada una de las tareas descritas anteriormente.

2.8.3.2. Planificación

La actividad de planificación consiste en identificar los diferentes procesos y actividades que serán realizados durante el desarrollo de la ontología, su planificación, el tiempo y recursos necesarios para su finalización. El objetivo es obtener una planificación concreta que permita guiar el desarrollo de la red ontológica[7].

A continuación se presenta en forma resumida las tareas establecidas por la metodología NeOn para llevar a cabo una planificación de forma adecuada:

Plantilla para el Documento de Especificación de Requisitos de la Ontología (ORSD)	
1	Propósito
	<i>El objetivo general de la ontología. Es decir, la función principal o rol que la ontología debe tener.</i>
2	Alcance
	<i>La cobertura general y el grado de detalle que la ontología debe tener.</i>
3	Lenguaje de Implementación
	<i>El lenguaje formal en el que la ontología debe estar implementada.</i>
4	Usuarios Finales Previstos
	<i>Los usuarios finales previstos para la ontología.</i>
5	Usos Previstos
	<i>Los usos previstos para la ontología.</i>
6	Requisitos
	a. Requisitos no Funcionales
	<i>Los requisitos o aspectos generales que la ontología debe satisfacer, incluyendo de manera opcional prioridades para cada requisito.</i>
	b. Requisitos Funcionales: Grupos de Preguntas de Competencia
	<i>Los requisitos relativos al contenido específico que la ontología debe satisfacer. Estos requisitos deben representarse en forma de preguntas de competencia y sus respuestas. Las preguntas de competencia deben agruparse, y se puede incluir de manera opcional prioridades para cada grupo y para cada pregunta de competencia.</i>
7	Pre-Glosario de Términos
	a. Términos de las Preguntas de Competencia
	<i>La lista de términos incluidos en las preguntas de competencia y sus frecuencias.</i>
	b. Términos de las Respuestas
	<i>La lista de términos incluidos en las respuestas y sus frecuencias.</i>
	c. Objetos
	<i>La lista de objetos incluidos tanto en las preguntas de competencia como en sus respuestas.</i>

Figura 2.9: Plantilla para el ORSD [3]

Tarea 1. Selección del modelo de ciclo de vida de la red ontológica: Establecer el modelo de ciclo de vida más apropiado para el desarrollo de la red ontológica.

Tarea 2. Selección del conjunto de escenarios: Seleccionar el conjunto de escenarios de la metodología NeOn (Véase Tabla 5) que se van a seguir durante el desarrollo de la red ontológica, para esto es necesario tener en cuenta tanto el documento el ORSD como los tipos de recursos que pueden ser reutilizados para el desarrollo.

Tarea 3. Actualización del plan inicial: Modificar si es necesario el plan inicial obtenido de la tarea anterior. Las modificaciones pueden ser: incluir o borrar procesos, actividades y fases o cambiar el orden y dependencias entre procesos y actividades.

Tarea 4. Establecer restricciones y asignar recursos: Incluir información relativa a la planificación temporal y la asignación de recursos humanos en cada uno de los procesos y actividades.

2.8.3.3. Reutilización de ontologías de dominio

Las ontologías de dominio proporcionan conocimiento específico sobre un dominio (medicina, farmacia, educación, etc.) y pueden ser útiles cuando se pretende desarrollar una ontología que represente el mismo dominio o similar.

El proceso de reutilización de ontologías de dominio tiene como objetivo encontrar y seleccionar una o varias ontologías relacionadas con el dominio de la ontología que se pretende construir para integrarlas a su desarrollo. Como resultado se obtendrá una red de ontologías extendida con las ontologías de dominio reutilizadas[7].

Las actividades a realizar para reutilizar ontologías de dominio planteadas en *Metodología Neon Aplicada a la Representación del Contexto*[7] se pueden resumir de la siguiente forma:

Actividad 1. Búsqueda de ontologías de dominio: Buscar en librerías, repositorios y registros ontologías de dominio candidatas que satisfagan las necesidades de la red de ontologías que se está desarrollando.

Actividad 2. Valoración de ontologías de dominio: Para valorar adecuadamente si una ontología de dominio candidata es o no útil para el proceso de reutilización, es necesario verificar los parámetros de alcance, propósito y cobertura de requerimientos funcionales y no funcionales. En el caso específico del propósito se verificará la similitud

entre las definiciones de los términos del ORSD y las establecidas en la ontología considerada.

La valoración de cada uno de los criterios antes mencionados se realiza en base a los siguientes parámetros:

- **Si-Totalmente (Si-T):** indica que la ontología candidata cumple de manera exacta con el criterio calificado
- **Si-Parcialmente (Si-P):** Indica que la ontología candidata cumple de manera parcial con el criterio calificado.
- **No (N):** Indica que la ontología candidata no cumple con el criterio calificado.
- **Desconocido (D):** Indica que la ontología candidata no tiene documentación suficiente para determinar si es válida o no para ser reutilizada dentro del criterio calificado.

La decisión de si una ontología es útil o no para el proceso de reutilización, es tomada en base a la siguiente heurística: Si el desarrollador contestó No a uno o todos los criterios de alcance similar, propósito similar o cobertura de requerimientos funcionales, entonces la ontología es considerada no útil y por lo tanto se elimina del conjunto de ontologías candidatas.

Actividad 3. Selección de las ontologías de dominio: El conjunto de ontologías candidatas es sometido a un proceso de selección que busca identificar las ontologías de dominio adecuadas para ser integradas a la red ontológica. Para ello se propone hacer uso de la tabla 2.4, en la cual se establecen cuatro dimensiones de criterios que deben ser calificados, el rango de valores a utilizar es de 0 a 3 que refieren a desconocido, bajo, medio y alto respectivamente.

Para establecer la puntuación total de cada una de las ontologías candidatas se hará uso de la ecuación 2.1.

$$Puntaje_i = \left(\sum_{j(+)} Valor_{Ti,j} \times \frac{Peso_j}{\sum_j Peso_j} \right) - \left(\sum_{j(-)} Valor_{Ti,j} \times \frac{Peso_j}{\sum_j Peso_j} \right) \quad (2.1)$$

Donde:

- **i:** es una ontología candidata particular.

Tabla 2.4: Plantilla que contiene los criterios de decisión para seleccionar ontologías de dominio. [7]

Criterio	Peso		Ontología 1	Ontología 2
Coste de reutilización				
Coste económico de reutilización	(-)	9		
Coste temporal de reutilización	(-)	7		
Esfuerzo de comprensión				
Calidad de la documentación	(+)	8		
Conocimiento externo disponible	(+)	7		
Claridad de código	(+)	8		
Esfuerzo de integración				
Adecuación para la extracción de conocimiento	(+)	9		
Adecuación del sistema de nombrado	(+)	5		
Adecuación del lenguaje de implementación	(-)	7		
Conflicto entre conocimiento representado	(+)	7		
Necesidad de términos puente	(-)	6		
Fiabilidad				
Disponibilidad de pruebas	(+)	8		
Evaluación anterior	(+)	8		
Reputación del equipo de desarrollo	(+)	8		
Fiabilidad del propósito	(+)	3		
Soporte práctico	(+)	7		

- j : es un criterio particular de los incluidos en la tabla 6, $j (+)$ representa criterios con peso positivo y $j (-)$ los criterios con peso negativo.
- $Valor_{Ti,j}$: es el valor para el criterio j en la ontología i .
- $Peso_j$: es el peso numérico asociado al criterio j .

Como resultado de este proceso se seleccionará la ontología cuyo puntaje total sea el más alto. En ciertos casos pueden ser seleccionadas más de una ontología dependiendo de su calidad y el aporte que puede tener a la red ontológica.

Actividad 4. Integración de las ontologías de dominio: El objetivo de esta actividad es integrar las ontologías de dominio seleccionadas en la red de ontologías que se está desarrollando. La integración puede realizarse reutilizando la ontología de dominio sin cambios, reutilizando la ontología de dominio con cambios significativos o mezclado varias ontologías que cubren el mismo dominio para obtener una nueva ontología.

2.8.4. Modelos de ciclo de vida para redes ontológicas

Un modelo de ciclo de vida para una red ontológica se define como un modelo para describir la forma en que debe realizarse un proyecto de red ontológica. En esta sección se describen dos modelos de ciclos de vida, los cuales se encuentran relacionados con el conjunto de escenarios descritos en la sección 2.8.2.

2.8.4.1. Modelo en cascada

Este modelo representa las etapas del desarrollo de una red ontológica como fases secuenciales, como una cascada, donde una etapa debe ser completada antes de iniciar con la siguiente. Para utilizar este modelo se asume que los requerimientos son completamente conocidos, sin ambigüedades e inalterables para iniciar con el desarrollo de la red ontológica[4].

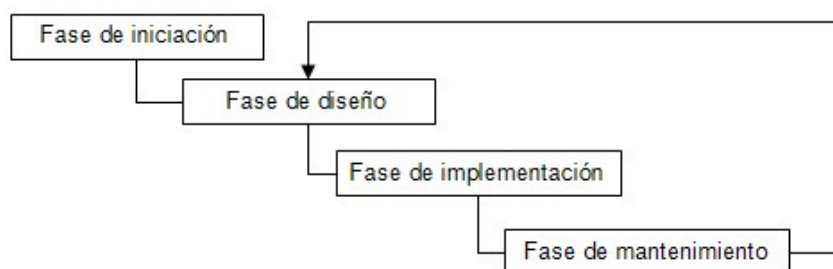


Figura 2.10: Ciclo de vida en cascada para redes ontológicas[4]

En la Figura 2.10, se presentan las actividades principales propuestas en el ciclo de vida en cascada:

1. **Fase de iniciación:** En esta fase debe definirse el ORSD y la planificación para el desarrollo de la red ontológica.
2. **Fase de diseño:** La salida de esta fase es un modelo informal y uno formal que satisface los requerimientos establecidos en el ORSD.
3. **Fase de implementación:** El modelo formal obtenido en la fase anterior es implementado en un lenguaje ontológico.
4. **Fase de mantenimiento:** Si durante el uso de la red ontológica se detectan errores o conocimiento faltante el grupo de desarrollo regresará a la fase de diseño para generar una nueva versión que corrija dichos inconvenientes.

2.8.4.2. Modelo iterativo incremental

Este modelo de ciclo de vida plantea realizar el desarrollo de una red de ontologías como un conjunto de iteraciones, donde cada iteración individual es similar a un proyecto que utiliza un modelo en cascada y entrega como resultado una red ontológica funcional que cumple un subconjunto específico de requerimientos[4], tal como se presenta en la Figura 2.11.

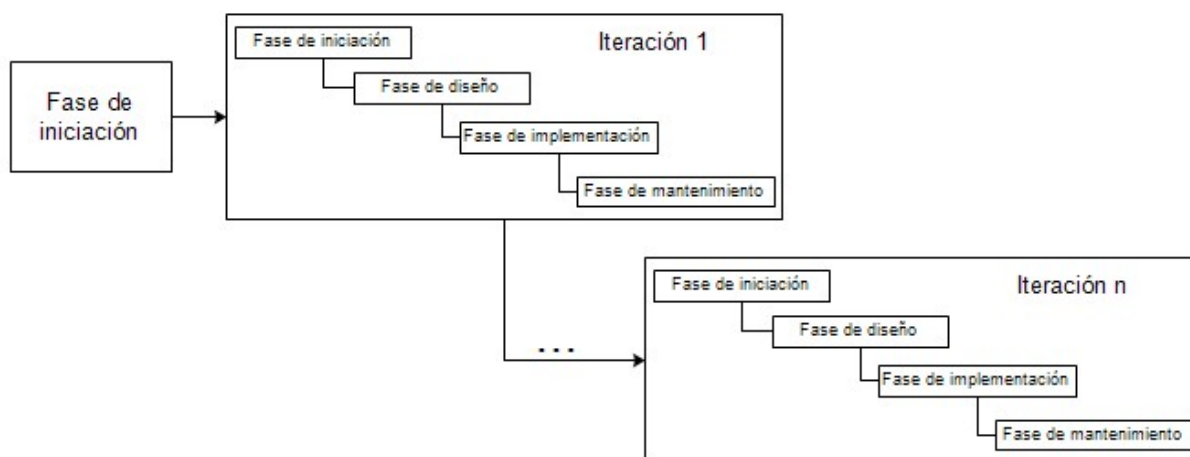


Figura 2.11: Esquema del modelo de ciclo de vida iterativo incremental[4]

Los principales beneficios de utilizar un ciclo de vida iterativo incremental para el desarrollo de una ontológica son:

- Identificar y mitigar los posibles riesgos tan pronto como sean identificados.



- Motivación del equipo de desarrollo por producir rápidamente una ontología adecuada.
- Pueden establecerse prioridades en el conjunto de requerimientos considerado para la iteración.
- El desarrollo puede ser adaptado a cambios en los requerimientos.
- La planificación de cada iteración puede ser adaptada en base a las experiencias de iteraciones previas.

Capítulo 3

Trabajos relacionados

3.1. Introducción

En la actualidad la web genera grandes cantidades de datos provenientes de distintas fuentes. En relación al dominio universitario, las diversas áreas existentes generan grandes cantidades de datos que son almacenados en repositorios particulares y representados haciendo uso de diversos formatos y modelos. Para mitigar estos problemas se plantea desarrollar modelos de datos que describan dicho dominio y su posterior publicación con tecnología semántica, utilizando los conceptos de ontologías y datos enlazados para lograr que la información sea debidamente tratada, abierta y compartida.

Las ventajas que ofrece la web semántica han inspirado el desarrollo de un gran número de investigaciones dirigidas a diversos dominios y que buscan dar solución a problemas del mundo real. Es por esto que, previo a la ejecución de un proyecto de investigación es necesario llevar a cabo una revisión del estado del arte en temas relacionados, con la finalidad de reutilizar conocimiento, experiencias y resultados obtenidos; y de esta manera ahorrar el tiempo y esfuerzo que tomaría el generar soluciones desde cero.

3.2. Modelos propuestos

El uso de tecnologías semánticas hace posible tener la descripción de los datos con un significado preciso y claro, es por esta razón que han sido aplicadas a diversos dominios, entre los cuales se encuentra el dominio universitario. Es por esto que, con la finalidad de tener un punto de partida y establecer información o recursos que podrían ser aplicables al presente trabajo, en este capítulo se han descrito cuatro trabajos relacionados, con su respectivo proceso de desarrollo y resultados obtenidos para cada caso.

3.2.1. Modelo ontológico para representar y publicar datos académicos

En [37] se plantea el desarrollo de una ontología que permite representar los datos de los planes de curso y la publicación de estos datos bajo las mejores prácticas de datos enlazados, con esto se pretende que las instituciones de educación superior puedan presentar una mayor visibilidad de su actividad académica mediante la publicación de datos abiertos, logrando poner a disposición de profesores y estudiantes información accesible y en formatos estándar que pueden ser empleados en proyectos futuros.

Para guiar el desarrollo de este proyecto se plantea emplear el ciclo de vida de datos enlazados propuesto por Villazón-Terrazas (Sección 2.7.3), en el cual se define la actividad de modelado del dominio de la información como segunda fase, para esto se hará uso de las actividades establecidas en la metodología NeOn (Sección 2.8). Por otra parte, el desarrollo del modelo ontológico para describir el dominio de planes de curso establece la reutilización de recursos ontológicos y no ontológicos, por lo cual, en base a la metodología NeOn se aplica la combinación de los escenarios:

- **Escenario 2. Reutilización y re-ingeniería de recursos no ontológicos:** Este tipo de recursos han sido buscados en sitios relacionados con el dominio del trabajo y la institución involucrada, en este caso se ha seleccionado la base de datos del Sistema de Planes. Posteriormente se requiere realizar un proceso de reingeniería donde dicha base de datos será transformada en una ontología. Para esto, se revisará su modelo relacional y evaluará cada uno de los conceptos para identificar cuales representarán una clase y qué atributos una propiedad.
- **Escenario 3. Reutilización de recursos ontológicos:** En este caso se ha reutilizado ontologías de dominio, enfocadas principalmente en representar conocimiento de un dominio específico. Dichos recursos han sido seleccionados e integrados a la ontología de planes de cursos, generando una red de ontologías que representan el dominio académico.

La ontología propuesta, cuyo modelo conceptual es presentado en la figura 3.1, esquematiza claramente los datos que representan el dominio de planificación de un curso, permitiendo que sean debidamente publicados. Adicionalmente el empleo de la metodología de NeOn permitió que se genere la construcción de una red de ontologías, en la cual se comparta el conocimiento de diferentes recursos ontológicos integrados en un solo modelo.

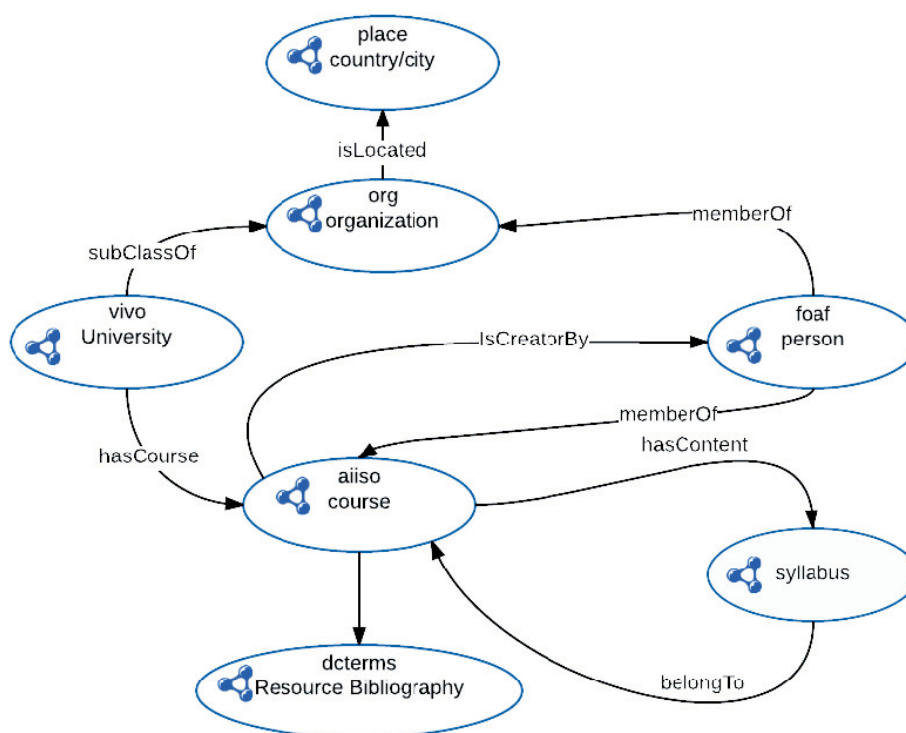


Figura 3.1: Modelo conceptual de la ontología de planes de curso

3.2.2. Linked Data Academic University (LDAU)

En un trabajo de investigación[5] realizado en la Universidad Nacional de Loja (UNL) se pretende estructurar semánticamente la información relacionada con la oferta académica de dicha universidad, con la finalidad de proveer al usuario un medio de consulta de información de datos relacionados con este dominio.

La metodología utilizada para la publicación de datos enlazados implica las etapas descritas a continuación:

1. **Identificación de fuentes de datos:** Se identifican los datos relacionados con la oferta académica de la UNL. En esta fase se establece:
 - a) **Términos del dominio:** Términos relacionados con el dominio en cuestión, por ejemplo: universidad, carrera, área, modalidad, nivel, período de oferta académica, etc.
 - b) **Datos que se van a entregar:** Se definen los datos que podrán ser recuperados por el usuario, por ejemplo: datos informativos generales de la UNL, carreras o programas de los diferentes niveles y modalidades ofrecidos por cada

área de la UNL, docentes por carrera, docentes asignados a una materia, entre otras.

- c) **Usuarios finales:** En este caso el proyecto está dirigido a estudiantes, docentes y público en general.
2. **Desarrollo de vocabularios:** En esta fase se recopilan los vocabularios u ontologías más utilizadas dentro del contexto establecido, con la finalidad de desarrollar un modelo ontológico específico para este caso.
 - a) **Identificar de clases y propiedades:** Se especifica un modelo conceptual en base a las clases y propiedades identificadas.
 - b) **Vocabularios consensuados:** Se definen los vocabularios que serán reutilizados dentro del contexto analizado, adicionalmente se establecen las clases y términos que serán reutilizados de cada uno de los vocabularios y los que serán creados como parte del vocabulario LDAU. En la figura 3.2 se presentan las clases y las relaciones definidas en el modelo ontológico obtenido como resultado.

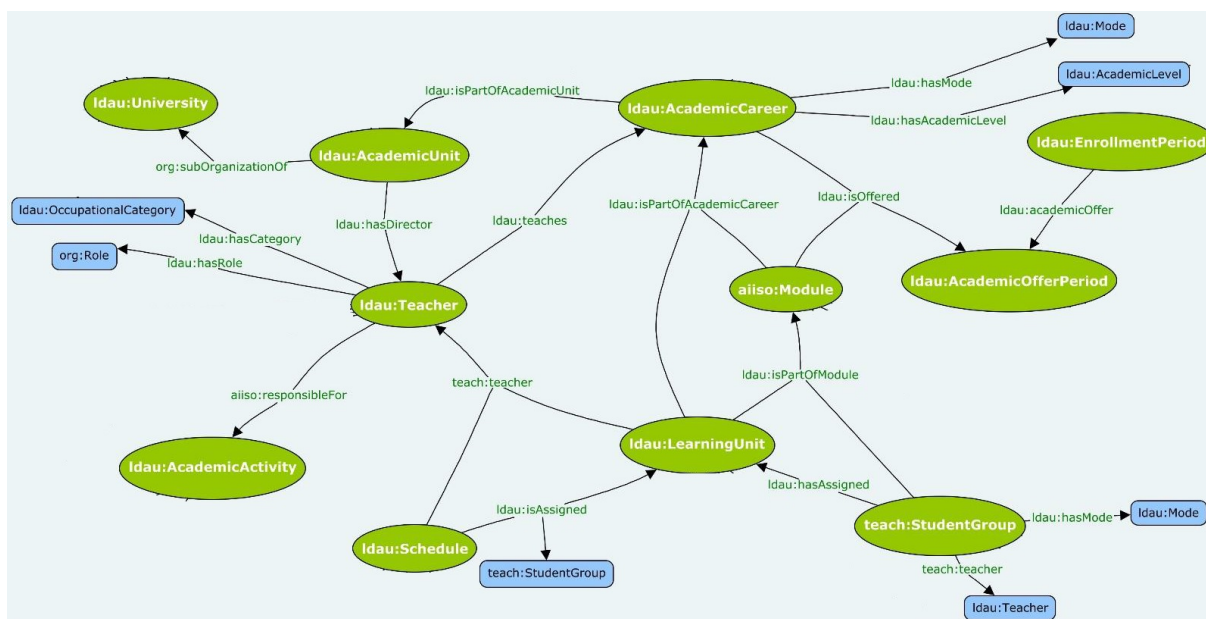


Figura 3.2: Clases y relaciones definidas en la ontología LDAU.

3. **Limpieza de datos:** Pretende detectar, corregir y eliminar errores e inconsistencias tales como:
- a) Datos incompletos.

- b) Errores en la estructura de un campo (fechas, números).
- c) Datos duplicados.

4. Generación de datos RDF:

- a) Definir el esquema de URIs tanto para el vocabulario como para las clases y propiedades.
 - b) Generación de tripletas RDF haciendo uso de herramientas especializadas.
 - c) Realizar un proceso de enriquecimiento de datos o enlazado, donde se busca identificar datos con los cuales se va a enlazar y descubrir relaciones entre conjuntos de datos.
5. **Publicación y explotación de datos RDF:** Una vez generadas las tripletas es necesario almacenarlas con la finalidad de realizar consultas sobre los datos RDF.
6. **Visualización:** Posterior a la publicación de los datos, es recomendable la utilización de una herramienta que permita visualizar los datos de manera amigable para los usuarios.

3.2.3. Ontology for Linked Open University Data (OLOUD)

OLOUD[38] es una ontología completamente generalizada que describe datos públicos disponibles en el dominio universitario y busca servir como base para la construcción de aplicaciones, por ejemplo aplicaciones móviles que permitan a los estudiantes organizar la hora y lugar de sus clases. Dicha ontología es capaz de responder preguntas relacionadas con unidades, empleados, asignaturas, cursos, horarios de clases, ubicaciones y eventos en una universidad.

OLOUD ha sido desarrollada mediante la reutilización de los vocabularios AIISO¹, AIISO-Roles², TEACH³, Event Ontology⁴, Time Ontology, W3C Basic Geo Vocabulary⁵, entre otros; y se han establecido dos módulos específicos:

- **Modulo principal OLOUD:** Provee clases y propiedades para describir la estructura organizacional de una universidad, asignaturas, cursos, eventos relacionados a la universidad, personas y sus publicaciones. La figura 3.3 muestra las clases principales y propiedades significativas de este módulo.

¹Academic Institution Internal Structure Ontology: <http://vocab.org/aiiso/>

²<http://vocab.org/aiiso-roles/>

³<http://linkedscience.org/teach/ns/>

⁴<http://motools.sourceforge.net/event/event.html>

⁵<http://www.w3.org/2003/01/geo/wgs84-pos>

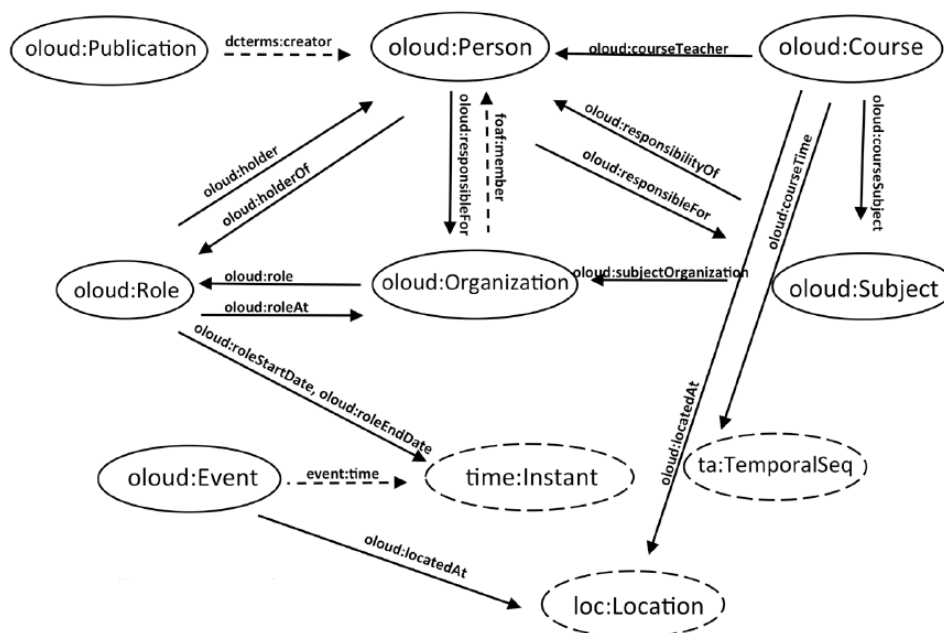


Figura 3.3: Clases y propiedades del módulo principal de OLOUD.

- **Módulo de localización/navegación en interiores:** Este módulo tiene por objetivo describir ubicaciones en interiores y características de navegación. Define clases para describir edificios y su estructura interna: pisos, habitaciones, pasillos y puntos de interés, tal como se muestra en la figura 3.4.

3.2.4. UH-Ontology

UH-Ontology es el modelo desarrollado por Yorbelis Rosell y sus colaboradores[39] que describe la ontología modelada para la gestión de datos heterogéneos en la Universidad de la Habana, que tiene por objetivo aportar consistencia y fiabilidad, para combatir el exceso de ambigüedad en la recuperación de información. UH-Ontology presenta las siguientes particularidades:

- **Alcance:** Modelar conceptos manejados respecto a personas, documentos (de texto, audio, video e imágenes), eventos o actividades planificadas, proyectos de software y espacios geográficos.
- **Usuarios potenciales:** Estudiantes, profesores, personal administrativo y directivos.
- **Usos específicos:** Búsqueda de recursos de información.

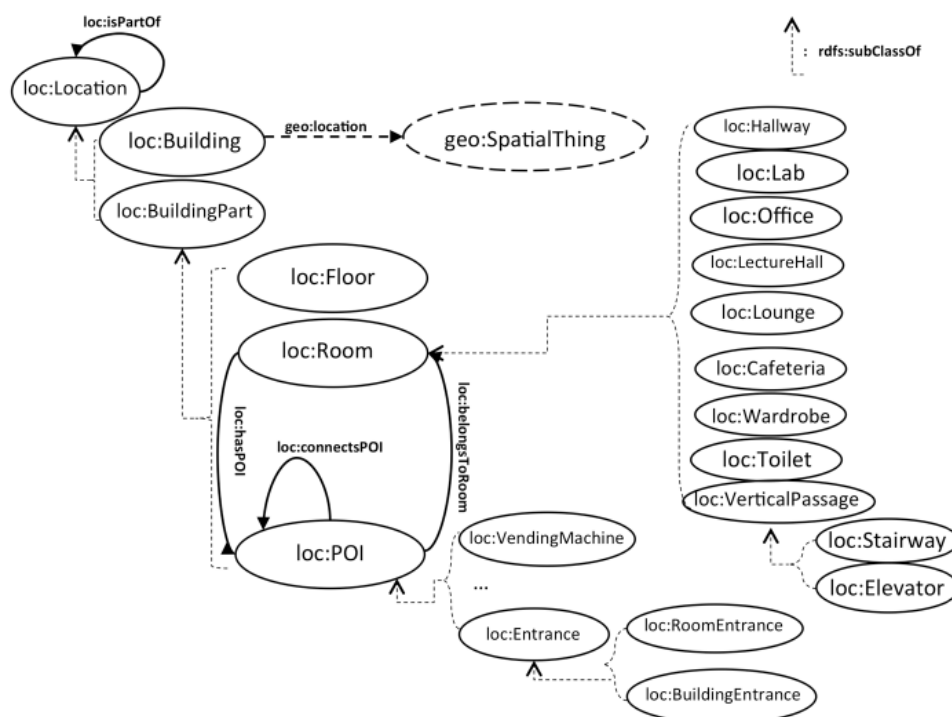


Figura 3.4: Estructura del módulo de localización/navegación en interiores.

Teniendo como premisa la reutilización de recursos existentes, UH-Ontology utiliza como base la ontología VIVO⁶. Adicionalmente se hizo uso de vocabularios como SKOS⁷, FOAF⁸, BIBO⁸ y Geontology para modelar particularidades de la universidad en cuestión. A continuación se presentan algunas de las clases principales definidas en el modelo resultante:

- **Agent:** Utilizada para describir cualquier agente que puede ser personas, organizaciones o grupos de cualquier tipo (Figura 3.5); por lo tanto, define las subclases:
- **Person:** Se centra en establecer los roles o funciones que pueden tener las personas en su relación con la UH y los recursos de información gestionados.
- **Organization:** Dedicada a identificar instituciones, organizaciones o grupos, sus diferentes estructuras y subordinaciones.
- **Document:** Contiene varias subclases que, a su vez, determinan otras relaciones subordinadas, permiten representar los diferentes recursos existentes y sus relaciones como “list of authors”, citedBy, reproducedIn, etc.

⁶<http://vivoweb.org/ontology/core#>

⁷<http://www.w3.org/2008/05/skos>

⁸<http://purl.org/ontology/bibo/>

- **EducationaTraining:** Refiere a los estudios universitarios posteriores al título de grado (maestría, doctorado, postdoctorado y cursos de especialización).
- **Event:** Clase creada para identificar los diferentes eventos o actividades planificadas de forma periódica.
- **Project:** Clase creada para identificar los diferentes eventos o actividades planificadas de forma periódica.

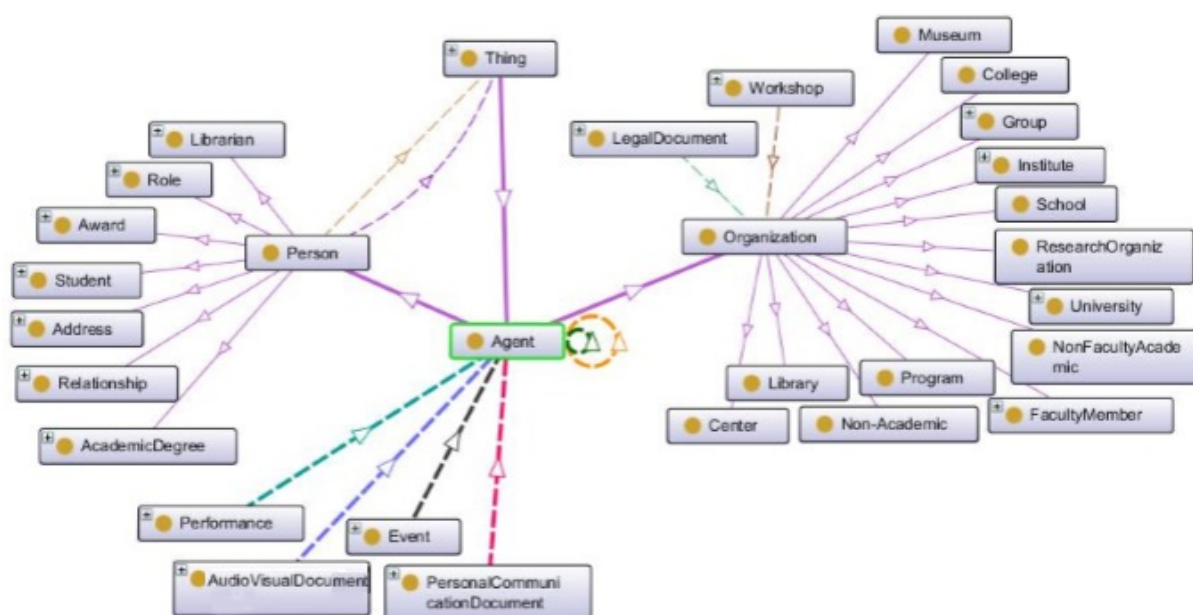


Figura 3.5: Clase Agent y sus relaciones con otras clases y subclases

3.2.5. Higher Education Reference Ontology (HERO)

Debido a que la mayoría de las ontologías son ontologías de aplicaciones, que no son reutilizables y son difíciles de vincular por ser demasiado específicas, Zemmouchi-Ghomari y Ghomari[40] plantean el desarrollo, siguiendo la metodología NeOn, de una ontología de referencia, que provea un modelo de conocimiento consensuado del dominio universitario y pueda ser considerado como base para la derivación de un modelo universitario más específico, dicha ontología es denominada HERO (Higher Education Reference Ontology) y su desarrollo es realizado siguiendo la metodología NeOn.

Debido a la complejidad del dominio y de la amplia cobertura de la ontología de referencia durante la selección de escenarios se opta por combinar el desarrollo desde cero con una estrategia de reutilización, siguiendo las siguientes fases:

- Durante la fase de especificación se reutilizan recursos no ontológicos aplicando una estrategia de búsqueda que consiste en centrarse en fuentes de información confiables (sitos web oficiales, asociaciones universitarias, informes académicos, etc.) y considerar las declaraciones relacionadas a la ontología como respuestas potenciales las preguntas de competencia.
- Durante la fase de conceptualización se buscarán y reutilizarán ontologías existentes en línea.

En base a la metodología NeOn los requerimientos ontológicos son identificados mediante el planteamiento de preguntas de competencia (81 en este caso), para posteriormente extraer la terminología a ser representada formalmente en la ontología mediante conceptos, atributos y relaciones. Parte del preglosario de términos obtenido, su clasificación y frecuencia se presenta en la figura 3.6.

Faculty, appointments and research area	Student and their life	Degrees and curriculum programs
Faculty member: 12	Service: 11	Associate degree: 10
Faculty: 36	Student: 74	Bachelor degree: 21
Research: 39	Administration	Course: 33
Teacher tenure: 11	Admission: 11	Credit: 29
Teacher: 12	Department: 14	Degree: 73
University: 17	Higher education institution: 13	Grade: 19
Finance	Organisation: 11	Grading: 11
Financial aid: 5	President: 10	Program: 34
Tuition: 5	Private: 10	

Figura 3.6: Preglosario de términos junto con sus respectivas frecuencias.

Para llevar a cabo la fase de conceptualización se integran técnicas de representación intermedia como diccionarios de datos, jerarquía de conceptos, arboles de clasificación de atributos y tabla de propiedades de objetos. En la figura 3.7 se presenta un la jerarquía existente entre los principales conceptos de HERO.

Posterior a la implementación es necesario evaluar la ontología con la finalidad de garantizar su correctitud, para esto se ha evaluado:

- La estructura lógica de la ontología con la finalidad de detectar inconsistencias y redundancias en las taxonomías de conceptos,

- El uso de la ontología mediante la evaluación por expertos del dominio y la técnica de preguntas de competencia que consiste en transformar el lenguaje natural de las preguntas de competencia a consultas SPARQL.
- El nivel de usabilidad que depende del de las anotaciones contenidas en la ontología y que la hacen más comprensible para el usuario final.

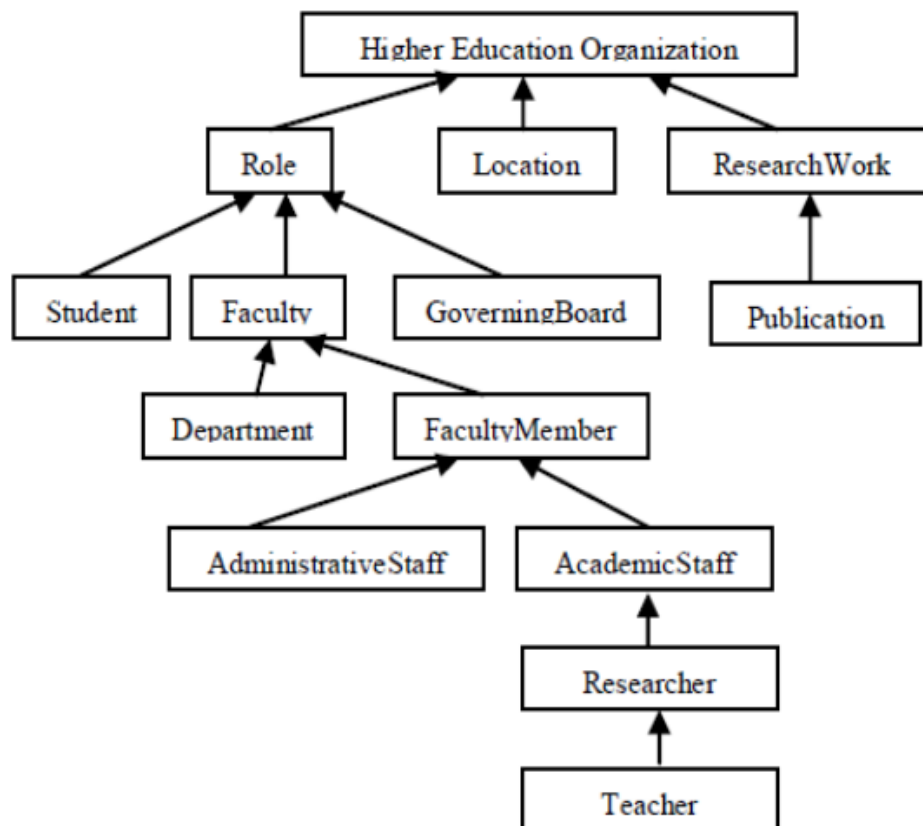


Figura 3.7: Jerarquía de conceptos clave de HERO

3.3. Resultados obtenidos

En base a los trabajos descritos los cuales se enfocan en describir distintas áreas del dominio universitario, se establece la tabla 3.1, que contiene, de forma resumida, los diversos modelos, su objetivo y las áreas consideradas en su alcance. Finalmente, se describe el aporte obtenido de cada uno de los trabajos analizados y que serán considerados durante la ejecución del presente proyecto de investigación.

Tabla 3.1: Resumen de trabajos relacionados

Modelo	Objetivo	Áreas consideradas
Para representar y publicar datos académicos	Desarrollar una ontología que permite representar los datos de los planes de curso y la publicación de estos datos bajo las mejores prácticas de datos enlazados.	Organización, personas, cursos, recursos bibliográficos y sílabos.
LDAU	Estructurar de forma semántica la información de la oferta académica de la UNL, con la finalidad de proveer al usuario un medio de consulta de información de datos relacionados con este dominio.	Unidades académicas, carreras, módulos, grupos de estudiantes, oferta académica, período de matrícula, profesores y horarios.
OLOUD	Describe datos públicos disponibles en el dominio universitario y busca servir como base para la construcción de aplicaciones.	Unidades, horarios de clases, empleados, asignaturas, cursos, ubicaciones y eventos.
UH-Ontology	Aportar consistencia y fiabilidad, para combatir el exceso de ambigüedad en la recuperación de información.	Personas, documentos, eventos o actividades planificadas, proyectos de software y espacios geográficos.
HERO	Proveer un modelo de conocimiento consensuado del dominio universitario que pueda ser considerado como base para derivar ontologías de dominio universitario más específicas.	Facultad, área de investigación, estudiantes, administración, finanzas, títulos y programas curriculares.

- **Modelo para representar y publicar datos académicos:** Describe la utilización del ciclo de vida para datos enlazados propuesto por Villazón – Terrazas y la metodología de NeOn para el modelado ontológico.
- **LDAU:** Este modelo describe la oferta académica de una universidad ecuatoriana, por lo cual incluye clases y propiedades útiles para el desarrollo del presente proyecto.
- **OLOUD:** Si bien el módulo principal de esta ontología contiene varios conceptos útiles, es el módulo creado para la localización/ubicación el que puede ser utilizado como base para la descripción de espacios físicos dentro de una universidad.

- **UH-Ontology:** Esta ontología tiene como prioridad modelar los diferentes tipos de documentos existentes en una universidad, lo cual es útil para el presente trabajo ya que se modelan trabajos de titulación.
- **HERO:** Al proveer un modelo ontológico de referencia del dominio universitario se convierte en una potencial ontología candidata que pudiera ser reutilizada en el presente proyecto.

En conjunto los trabajos mencionados aportan guías sobre el desarrollo de proyectos relacionados con la publicación de datos enlazados, además de indicar vocabularios que pueden ser reutilizados para el modelamiento ontológico de los diversos ámbitos requeridos.

Capítulo 4

Proceso de integración de datos

En este capítulo se describe la ejecución de cada una de las actividades planteadas en el ciclo de vida para datos enlazados, propuesto por Villazón-Terrazas (sección 2.7.3), llevadas a cabo con la finalidad de integrar los datos existentes en diferentes fuentes heterogéneas en un único recurso, sobre el cual, posteriormente se permitirá la ejecución de búsquedas.

Para el desarrollo adecuado de este proceso es necesario el uso de herramientas que permitan facilitar o automatizar las tareas requeridas, permitiendo un ahorro de tiempo y esfuerzo. A continuación, se presenta un breve resumen de las herramientas utilizadas en el presente proyecto.

- **Protegé v5.2:** Para la implementación del modelo ontológico desarrollado durante la etapa de modelado.
- **Ubuntu 14.6 LTS:** Sistema operativo base.
- **LOD-GF¹:** Framework desarrollado en la Universidad de Cuenca, basado en PDI² que brinda un entorno unificado para el soporte de cada una de las fases de la metodología de publicación de LOD[41]. Los plugins utilizados se describen a continuación:
 - **Data Precatching:** Es un plugin de almacenamiento temporal que permite que los datos puedan ser manipulados posteriormente.

¹LOD-GF: Linked Open Data - Generation Framework. *Disponible en:* <https://github.com/ucuenca/lodplatform>

²PDI: Pentaho Data Integration o también conocido como kettle. Es un diseñador gráfico de transformaciones y trabajos de ETL (extracción, transformación y carga).

- **Get Properties:** Utilizado en la etapa de modelamiento, permite cargar los vocabularios de ontologías tanto de las que se encuentran disponibles en la web, como localmente.
- **Ontology Mapping:** Permite describir los datos semánticamente empleando los vocabularios de las ontologías previamente cargadas, para ello, vincula los recursos con determinado vocabulario semántico.
- **Fuseki Loader:** Se utiliza en la etapa de publicación y permite configurar los parámetros básicos para el despliegue del triplestore³ Fuseki.
- **ELDA Loader:** Durante la etapa de Explotación este plugin ofrece la posibilidad de generar una página de descripción de los recursos, en la cual se puede consumir y navegar por la información generada en formato RDF de una forma amigable para el usuario.

4.1. Especificación

Al iniciar el desarrollo del proyecto de datos enlazados es necesario delimitar su alcance, con el objetivo de servir como base para el desarrollo de las actividades posteriores. Adicionalmente, dentro de esta actividad se lleva a cabo el análisis de las fuentes de datos para establecer los recursos de los que serán extraídos los datos necesarios; y la definición de la URI que será utilizada par identificar los recursos ontológicos.

4.1.1. Definición del alcance total

Con la finalidad de identificar adecuadamente las fuentes de datos requeridas para el proyecto se ha determinado el alcance general para el proyecto el cual limita las áreas y datos que serán cubiertos. Dicho alcance se encuentra definido en base a:

- **Estructura organizacional:** Organigrama⁴ que describe la estructura orgánica funcional y estatuto oficial⁵ de la Universidad de Cuenca.
- **Estructura académica:** En este caso se considera la información descrita a continuación:
 - **Organización académica:** Establece el siguiente orden de jerarquía:

³Triplestore o Tienda RDF: Base de datos diseñada especialmente para el almacenamiento y recuperación de tripletas mediante consultas semánticas.

⁴<https://www.ucuenca.edu.ec/sobre-uc/transparencia/info-administrativa/estructura-org-funcional#2017>

⁵https://www.ucuenca.edu.ec/images/Documentos_PDF/ESTATUTO_APROBADO_CES.18-DICIEMBRE-2013.pdf



- Facultad
- Carrera
- Plan de carrera
- Malla curricular
- Asignatura
- Sílabo
- **Asignaturas:** Especifica información propia de la asignatura e información fundamental para la formación de la malla curricular:
 - Código
 - Nombre
 - Número de créditos
 - Prerrequisitos
 - Correquisitos
 - Asignaturas equivalentes
- **Sílabos:** Al ser instrumentos de planificación es importante considerar información sobre el contenido de la asignatura y los métodos de evaluación a aplicar.
- **Estudiantes:** Considera los siguientes atributos contenidos en la ficha socio-económica:
 - **Datos personales:** Cédula o pasaporte, apellidos, nombres, nacionalidad, lugar y fecha de nacimiento, estado civil, género, tipo de sangre, lateralidad, lugar de domicilio, teléfono convencional y celular, correo electrónico personal e institucional, etnia con la que se identifica, tipo de discapacidad, porcentaje de discapacidad.
 - **Educación secundaria:** Colegio de graduación, título de bachiller, año de graduación.
 - **Educación universitaria:** Ha estudiado en otra universidad, nombre de la universidad, ha cursado una carrera diferente a la actual en la Universidad de Cuenca.
 - **Vivienda familiar:** Tenencia de la vivienda, material de construcción, zona de residencia, tipo de vivienda, ubicación, dirección, teléfono, pago mensual por arriendo y/o hipoteca.
 - **Bienes familiares:** Número de propiedades no renteras, renteras y vacacionales, número de líneas telefónicas, número de vehículos, valor acumulado por avalúo de vehículos, dispone de TVCable o similares.
 - **Grupo familiar:** Información personal del jefe de familia, número de integrantes, número de miembros que estudian en niveles diferentes al superior,

número de hijos menores de 6 años, es carga familiar de un servidor de la universidad.

- **Ingresos y egresos del grupo familiar:** Descripción y valor de los diferentes ingresos y egresos mensuales.
- **Oferta académica:** Incluye información sobre:
 - Período académico
 - Asignatura ofertada
 - Docente responsable
 - Horario de clases
 - Información sobre el paralelo
 - Espacio físico en el que se dicta clases.
- **Matrículas:** Constituye un vínculo entre un estudiante y una oferta académica; además permite establecer el registro académico correspondiente.
- **Registro académico:** Contiene información referente a cada una de las asignaturas en la que un estudiante se ha matriculado:
 - Nota final obtenida
 - Estado de aprobación (aprobado, reprobado, suspenso).
 - Forma de aprobación (escolaridad, homologación).
- **Estudiantes graduados:** Implica el registro de información propia del acta de graduación:
 - Número de refrendación del acta de grado.
 - Fechas de inicio de estudios, egreso y graduación
 - Mecanismo de titulación
 - Nota de graduación
- **Trabajos de titulación:** Requiere el registro de la siguiente información:
 - Título
 - Autor o autores
 - Director
 - Calificación obtenida
 - URL en la que se encuentra disponible.

Para una mejor comprensión del alcance académico del proyecto, la información descrita anteriormente ha sido organizada en forma jerárquica, dando como resultado el organigrama presentado en la figura 4.1.

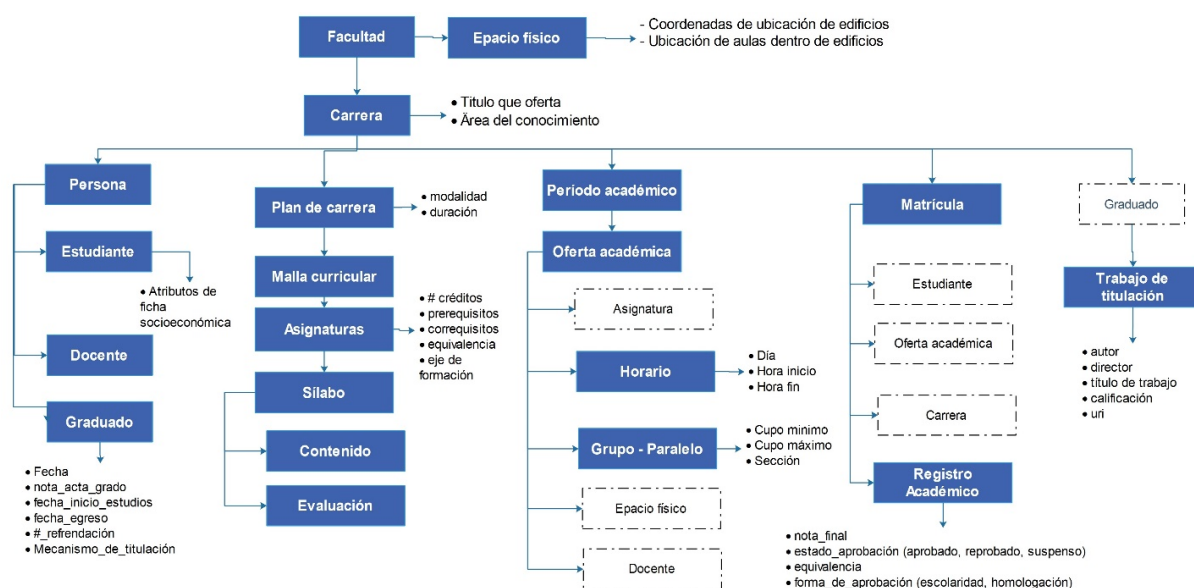


Figura 4.1: Organigrama que describe el alcance académico del proyecto.

4.1.2. Identificación y análisis de fuentes de datos

El objetivo de esta etapa es identificar y analizar fuentes de datos potenciales (bases de datos Oracle, AS400) que posteriormente serán convertidas en tripletas RDF. Para tal fin, se han analizado los esquemas disponibles en las bases de datos de la Universidad de Cuenca. En la tabla 4.1 se presenta el listado de las fuentes de interés identificadas.

Tabla 4.1: Fuentes de datos obtenidas como resultado del proceso de identificación y análisis

Fuente	Contenido	Tipo
Organigrama de la Universidad de Cuenca	Organismos que forman la Universidad de Cuenca y su jerarquía	PDF
ADMINUC	ESPACIOS.FISICOS PERSONAS.NATURALES TÍTULOS.PERSONAS INSTITUCIONES.EDUCATIVAS UBICACIONES TITULO PERIODOS.LECTIVOS PERSONAS TELEFONOS.PERSONAS BLOQUES	BD (ORACLE)



Fuente	Contenido	Tipo
ACADEMICO	CARRERAS MATRICULAS HORARIOS_OFERTAS SGA_MATRICULAS SGA_CARRERAS_ESTUDIANTES VST_ACTAS SGA_DATOS_GRADUADOS SGA_TITULOS_CARRERAS SGA_AREAS_CONOCIMIENTO SGA_PLANES_CARRERAS SGA_PLANES_MODALIDADES SGA_MODALIDADES SGA_MALLAS SGA_ASIGNATURAS SGA_ASIGNATURAS_MALLAS SGA_EQUIPARACIONES SGA_PRERREQUISITOS SGA_EJES_FORMACION SGA_DATOS_GRADUADOS SGA_ACTA_ESTUDIANTES SGA_OFERTA_ASIGNATURAS SGA_OFERTA_DOCENTES SGA_MATRICULAS_ASIGNATURA SGA_TIPOS_INSCRIPCIONES	BD (ORACLE)
GSOCIOECONOMICA	MV_FICHASSE	BD (ORACLE)
GHUMANA	SERVIDORES	BD (AS 400)
PRERREQUISITO	PARAACADEMICO	BD (ORACLE)
SILABO	SGS_SILABO SGS_CONTENIDO_CABECERA SGS_CONTENIDO_DETALLE SGS_SUBCAPITULOS SGS_METODOS_EVALUACION SGS_METODOS SGS_SUBCAPITULOS	BD (ORACLE)

4.1.3. Diseño de URI

Para la definición de la URI base se ha considerado el repositorio en el cual estará alojada tanto la ontología como sus instancias, el resultado de esta actividad se presenta en la tabla 4.2.

Tabla 4.2: Diseño de URIs a utilizar en el proyecto de datos enlazados

URI	Tipo	Descripción
http://ucuenca.edu.ec/	Base	Uri base para el modelo
http://ucuenca.edu.ec/ontology/{Concepto—propiedad}	TBox	Uri a utilizar para la creación de conceptos o propiedades. <ul style="list-style-type: none">■ Para conceptos la primera letra debe ser mayúscula.■ Para propiedades la primera letra debe ser minúscula.
http://ucuenca.edu.ec/resource/{resourcetype}/{resourcename}	ABox	Para definir instancias se adjuntará el tipo de recurso y finalmente un identificador único.

4.2. Modelo ontológico (NeOn)

El desarrollo de la red de ontologías requerida para describir el dominio de la Universidad de Cuenca, se realizará mediante las etapas o fases establecidas en la metodología NeOn (Sección 2.8), la cual incluye especificación de requerimientos, selección de ciclo de vida y escenarios, conceptualización del conocimiento, reutilización de recursos, formalización e implementación.

Inicialmente se lleva a cabo la fase de especificación de requerimientos con la que se obtiene el llamado Documento de Especificación de Requerimientos Ontológicos (ORSD), que es utilizado como guía a lo largo del desarrollo del modelo ontológico, teniendo en cuenta que este debe satisfacer los requisitos establecidos en dicho documento. Posteriormente se realiza una búsqueda de información y recursos de conocimiento relacionados con el dominio que se pretende modelar, en base a los resultados obtenidos de este proceso se seleccionan el o los escenarios adecuados para el desarrollo ontológico y se realizan las actividades establecidas en cada uno de ellos.

4.2.1. Especificación de requerimientos

En esta sección se describe el desarrollo de la guía propuesta en la metodología NeOn para cubrir la actividad de especificación de requerimientos ontológicos descrita en la sección 2.8.3.1. El objetivo de esta fase es elaborar el ORSD que contiene los requisitos que permiten modelar la estructura organizacional y académica de la Universidad de Cuenca, dicho documento ha sido descrito haciendo uso de la plantilla presentada en la figura 2.9. El documento completo puede ser observado en el anexo A.

En la tabla 4.3, se muestran las cinco primeras secciones del ORSD donde se describen los resultados de las tres primeras tareas establecidas para el proceso de especificación de requerimientos. El propósito y alcance del modelo ontológico fueron establecidos en base al alcance general establecido para el proyecto de datos enlazados en la fase de especificación del ciclo de vida propuesto por Villazón – Terrazas.

En cuanto a la identificación de requerimientos, en el caso específico de los requisitos funcionales, se ha hecho uso de la técnica de preguntas de competencia, las cuales fueron establecidas de tal manera que permitan describir los organigramas planteados como alcance total del proyecto (sección 4.1.1), para esto y para la obtención de las respuestas a cada una de las preguntas, se ha consultado información en la página web oficial⁶, sistemas existentes (SIUC⁷, ESIUC⁸), fuentes de datos e información disponible en las áreas relacionadas de la Universidad de Cuenca. Como resultado se obtuvo un total de 118 preguntas que la ontología construida debe ser capaz de responder.

En la tabla 4.4 se presenta, a modo de ejemplo, parte del grupo de preguntas de competencia que describen la organización académica definida como parte del alcance total en la figura 4.1.

Siguiendo con el proceso de especificación de requerimientos, las preguntas de competencia han sido agrupadas y priorizadas respecto al tipo de objeto al que hacen referencia, definiendo los siguientes grupos:

1. Organización administrativa
2. Organización académica
3. Espacios físicos

⁶<https://www.ucuenca.edu.ec/>

⁷Sistema Integrado de la Universidad de Cuenca

⁸Sistema Integrado de la Universidad de Cuenca para Estudiantes

Tabla 4.3: Documento de especificación de requerimientos: Secciones 1 a 5

1	Propósito
Proveer un modelo genérico de conocimiento que permita describir y presentar conceptos y relaciones existentes en el dominio universitario. Dicho modelo debe ser capaz de responder a consultas referentes a la estructura organizacional y académica de la Universidad de Cuenca. Adicionalmente el modelo obtenido debe almacenar la información necesaria para ser utilizado posteriormente en un sistema de búsqueda semántica.	
2	Alcance
La ontología denominada UCuenca-Ontology plantea modelar la estructura organizacional y académica de la Universidad de Cuenca, para lo cual es necesario el almacenamiento de información de diversa índole dentro del contexto universitario. Dicha información se plantea en base a:	
<ul style="list-style-type: none">▪ <i>Estructura organizacional:</i> Organigrama de la estructura orgánica funcional y Estatuto de la Universidad de Cuenca▪ <i>Estructura académica:</i> Organigrama planteado posterior al estudio de los sistemas y fuentes de datos disponibles en la Universidad de Cuenca.(figura 4.1)	
3	Lenguaje de implementación
El lenguaje a utilizar para la implementación de la ontología es OWL/XML	
4	Usuarios finales previstos
<ul style="list-style-type: none">▪ Usuario 1: Autoridades▪ Usuario 2: Personal docente y administrativo▪ Usuario 3: Estudiantes	
5	Usos previstos
<ul style="list-style-type: none">▪ Uso 1: Consultar información relacionada a la estructura organizacional▪ Uso 2: Consultar información relacionada a la estructura académica▪ Uso 3: Permitir conexión con modelos universitarios heterogéneos de áreas complementarias.	

4. Trabajos de titulación
5. Información personal
6. Información socioeconómica

Como tarea final se ha obtenido el pre-glosario de términos, considerando tanto las preguntas de competencia como sus respuestas, el objetivo principal es extraer nombres de entidades u ocurrencias, es decir, objetos del dominio de estudio[42]. Para hacer más sencilla y natural la representación de conceptos, atributos, propiedades e instancias el pre-glosario ha sido dividido en tres categorías: nombres, adjetivos y verbos.

Tabla 4.4: Parte del grupo de preguntas de competencia sobre organización académica

b. Requerimientos funcionales – Preguntas de Competencia		
ID.	Pregunta de competencia (CQ)	Respuesta
PC10	¿Cuál es el plan de carrera actual de la carrera X?	Plan de carrera 1
PC11	¿Qué título universitario otorga la carrera X?	Título 1
PC13	¿Cuál es la modalidad de estudios de la carrera X?	Presencial /semipresencial /Distancia
PC16	¿Cuál es la malla curricular de la carrera X?	Malla curricular 1
PC17	¿Cuáles son las asignaturas de la malla curricular X?	Asignatura 1, Asignatura 2,...
PC19	¿Cuál es el código de la asignatura X?	2262
PC23	¿Cuáles son los prerrequisitos de la asignatura X?	Asignatura 1, Asignatura 2,...
PC24	¿Cuáles son los correquisitos de la asignatura X?	Asignatura 1, Asignatura 2,...
PC25	¿A qué asignaturas es equivalente la asignatura X?	Asignatura 3 Asignatura 4 + Asignatura 5

4.2.2. Selección del ciclo de vida y escenarios

Una vez concluido el proceso de especificación de requerimientos ontológicos, el siguiente paso es establecer el ciclo de vida y escenarios que serán aplicados al desarrollo de la red ontológica. Esto permitirá definir las actividades que deben ser realizadas y con ello planificar de mejor manera el tiempo y recursos necesarios para la culminación del

proyecto.

Previo a seleccionar el ciclo de vida a utilizar, es necesario realizar una búsqueda rápida de recursos ontológicos a reutilizar, para esto, considerando los términos incluidos en el pre-glosario definido en el ORSD (anexo A), se han definido los conceptos, propiedades y atributos que requieren ser representados. Además, debido a que la mayoría de recursos ontológicos hacen uso del idioma inglés, se ha visto en la necesidad de traducir los términos a este lenguaje. En la tabla 4.5 se presenta parte de los conceptos definidos con su respectiva traducción.

Tabla 4.5: Parte de los conceptos a incluir en el modelo ontológico

Concepto	
Español	Inglés
Asignatura	Course, Subject
Carrera	Academic career
Departamento (Dirección / Unidad / Coordinación)	Department
Docente	Teacher
Estudiante	Student
Facultad	Faculty
Oferta académica	Academic offer
Organismo de investigación (Dirección / departamento / grupo de investigación)	Research organization
Plan de carrera	Curriculum
Malla curricular	Curriculum map
Universidad	University

Con la finalidad de facilitar la selección del ciclo de vida y escenarios adecuados para el desarrollo de la red ontológica, en la tabla 4.6 se presenta un conjunto de preguntas y sus respectivas respuestas para el presente proyecto. Para responder correctamente cada una de las preguntas, previamente se debe considerar que:

- El escenario 1 contiene las actividades principales que deben ser realizadas en cualquier desarrollo ontológico y debe ser combinado con los otros escenarios.
- Luego de realizar la actividad de especificación de requerimientos, se recomienda llevar a cabo una búsqueda rápida de recursos de conocimiento candidatos (recursos ontológicos y no ontológicos; y patrones de diseño) que puedan ser reutilizados en el desarrollo, utilizando como entrada los términos incluidos en el ORSD[4].

Tabla 4.6: Preguntas para establecer ciclo de vida y escenarios a utilizar

Define		Pregunta	Si	No
CICLO DE VIDA		¿Los requerimientos ontológicos son completamente conocidos, sin ambigüedades e inalterables al iniciar el desarrollo de la red ontológica?	X	
ESCENARIO	1	De la especificación a la implementación	X	
	2	¿Tiene planeado utilizar recursos no ontológicos como tesauros, bases de datos, etc. en el desarrollo de la red ontológica?		X
	3	¿Tiene planeado utilizar recursos ontológicos existentes en el desarrollo de la red ontológica?	X	
	4	¿Tiene planeado utilizar y modificar recursos ontológicos existentes en el desarrollo de la red ontológica?		X
	5	¿Tiene planteado utilizar y fusionar un conjunto de recursos ontológicos existentes en el desarrollo de la red ontológica?	X	
	6	¿Tiene planeado utilizar, fusionar y modificar un conjunto de recursos ontológicos existentes en el desarrollo de la red ontológica?		X
	7	¿Tiene planeado utilizar patrones de diseño en el desarrollo de la red ontológica?		X
	8	¿Tiene planeado reestructurar la red ontológica?	X	
	9	¿Tiene planeado desarrollar la red ontológica en diferentes lenguajes naturales?		X

En base a las respuestas establecidas, para el desarrollo de la red ontológica UCuenca-Ontology se plantea utilizar:

- **Ciclo de vida en cascada:** Los requerimientos son completamente conocidos y no cambiarán durante el desarrollo de la red ontológica.
- **Escenario 1. De la especificación a la implementación:** Debe ser utilizado en todo desarrollo ontológico.
- **Escenario 5. Reutilización y fusión de recursos ontológicos:** Si bien existen diversos vocabularios que pueden ser reutilizados para modelar las áreas existentes en el dominio universitario, estos deben ser fusionados para dar lugar a una red

ontológica adecuada. Cabe mencionar que este escenario contiene las actividades reutilización del escenario 3, es por ello que la pregunta correspondiente a dicho escenario se encuentra marcada.

- **Escenario 8: Re-estructuración de recursos ontológicos:** El proceso de alineamiento realizado como parte de la fusión de ontologías, no tiene sentido sin un proceso de re-estructuración de los recursos ontológicos a ser alineados[43], ya que pueden darse situaciones como:
 - Los recursos seleccionados contienen información irrelevante para el dominio en cuestión.
 - Algunos conceptos relevantes para el dominio no se contemplan en los recursos seleccionados.

4.2.3. Reutilización ontologías de dominio

Con el objetivo de encontrar y seleccionar una o varias ontologías relacionadas al dominio de cada una de las áreas existentes en una universidad e integrarlas a la red ontológica, se llevará a cabo las actividades de reutilización de ontologías de dominio descritas en la sección 2.8.3.3

El resultado de la búsqueda de ontologías de dominio candidatas realizada en base a los términos definidos en el ORSD, organizadas por grupo son presentadas en la tabla 4.7, donde puede observarse, que no se ha considerado el grupo "Información socioeconómica", esto se debe a que una vez finalizado el proceso de búsqueda de ontologías para este dominio no se obtuvo resultado alguno, por lo tanto, los términos relacionados a este grupo serán agregados a la red ontológica posteriormente durante el proceso de re-estructuración.

Continuando con el proceso, es necesario valorar el grupo de ontologías candidatas, para ello, en lo relacionado a los parámetros de alcance y propósito similar, estos se resumen en la tabla 4.8; mientras que para determinar la cobertura de requerimientos funcionales, cada una de las ontologías candidatas fue analizada para determinar la coincidencia entre los conceptos definidos y los términos del ORSD.

⁹<http://swat.cse.lehigh.edu/onto/univ-bench.owl>

¹⁰<https://www.w3.org/TR/vocab-org/>

¹¹<https://old.datahub.io/dataset/higher-education-reference-ontology>

¹²http://data.opendiscoveryspace.eu/lom_ontology_ods.owl

Tabla 4.7: Ontologías candidatas por grupo

Grupo	Ontologías candidatas
Organización administrativa	<ul style="list-style-type: none">■ Univ-Bench⁹■ AIISO■ ORG (Core Organization Ontology) ¹⁰■ HERO (Higher Education Reference Ontology)¹¹
Organización académica	<ul style="list-style-type: none">■ HERO■ LOM(Learning Object Metadata Ontology)¹²■ TEACH (Teaching Core Vocabulary Specification)
Espacios físicos	<ul style="list-style-type: none">■ WGS84 Basic Geo■ VIVO-ISTF
Trabajos de titulación	<ul style="list-style-type: none">■ BIBO (Bibliographic Ontology)
Información personal	<ul style="list-style-type: none">■ FOAF■ VCARD

Tabla 4.8: Propósito y alcance de ontologías candidatas para los grupos organización administrativa y organización académica

Vocabulario	Propósito	Alcance
Univ-Bench	Describir universidades y sus departamentos, además de las actividades que ocurren en ellos.	Ontología de dominio universitario, cubre el área organizacional, personas, publicaciones, trabajos, etc.
AIISO	Describir la estructura organizacional interna de una institución académica.	Cubre básicamente la estructura organizacional, sin incluir a personas.
ORG	Apoyar la publicación de datos enlazados sobre la información de organizaciones en general.	Términos básicos mínimos para representar estructura organizativa, personas, roles, ubicación e historia organizacional.
HERO	Ser relevante o al menos conveniente para describir cualquier universidad.	Describe varios aspectos del dominio universitario como la estructura organizativa, administración, personal, roles, ingresos, etc.
LOM	Utilizado para exponer IEEE LOM, un estándar de metadatos para contenidos educativos.	Describe objetos de aprendizaje, requerimientos técnicos, duración, etc.
TEACH	Proporciona términos que permiten a los maestros relacionar las cosas en sus cursos.	Describe cursos, asignaturas, materiales, estudiantes, y las relaciones entre ellos.

En la tabla 4.9 se presenta el resultado de la valoración de las ontologías candidatas para el dominio de la organización administrativa, como resultado ORG y HERO son descartadas debido a:

- ORG no tiene un propósito similar ya que permite modelar cualquier tipo de organización, adicionalmente al contener términos básicos mínimos no cubre los requerimientos funcionales necesarios.
- HERO difiere en el alcance ya que se centra mayormente en describir la parte académica de una universidad, por lo tanto la cobertura de requerimientos funcionales es mínima.

Por otra parte, en la tabla 4.10, se valoran las ontologías candidatas para el dominio de la organización académica, descartando TEACH y LOM en base a lo siguiente:

- TEACH no tiene un alcance similar ya que se centra únicamente en describir cursos, que son solamente una parte del dominio de la organización académica considerada para este proyecto.
- LOM tiene un propósito diferente, ya que está orientado al estándar IEEE LOM.

Tabla 4.9: Resultado de la valoración de ontologías candidatas para el grupo organización administrativa

	Univ-Bench	HERO	AIISO	ORG
Alcance similar	SI-P	N	SI-P	SI-P
Propósito similar	SI-P	SI-P	SI-P	N
Cobertura de requerimientos no funcionales	-	-	-	-
Cobertura de requerimientos funcionales	SI-P	N	SI-P	N

Tabla 4.10: Resultado de la valoración de ontologías candidatas para el grupo organización académica

	HERO	TEACH	LOM
Alcance similar	SI-P	N	SI-P
Propósito similar	SI-P	SI-P	N
Cobertura de requerimientos no funcionales	-	-	-
Cobertura de requerimientos funcionales	SI-P	SI-P	SI-P

Finalmente, debido a que luego de la valoración ontológica para el grupo de la organización administrativa, se han obtenido dos ontologías se requiere realizar un proceso de selección con la finalidad de identificar la ontología de dominio adecuada para formar parte de la red ontológica UCUenca-Ontology.

En la tabla 4.11 se muestran las puntuaciones obtenidas en cada uno de los criterios considerados, las cuales se establecieron luego de analizar la ontología (clases y propiedades a reutilizar, claridad del código, representación del conocimiento, etc.) e información disponible en la página oficial. Como puede observarse luego de la aplicación de la ecuación 2.1, descrita en la sección 2.8.3.3, la ontología Univ-Bench ha obtenido una puntuación total mayor, por lo cual es seleccionada para formar parte de la red ontológica.

Tabla 4.11: Selección de ontologías para la organización administrativa

Criterio	Peso		Univ-Bench	AIISO
Coste de reutilización				
Coste económico de reutilización	(-)	9	1	1
Coste temporal de reutilización	(-)	7	1	2
Esfuerzo de comprensión				
Calidad de la documentación	(+)	8	2	3
Conocimiento externo disponible	(+)	7	2	1
Claridad de código	(+)	8	3	3
Esfuerzo de integración				
Adecuación para la extracción de conocimiento	(+)	9	3	2
Adecuación del sistema de nombrado	(+)	5	3	1
Adecuación del lenguaje de implementación	(-)	7	3	3
Conflicto entre conocimiento representado	(+)	7	1	1
Necesidad de términos puente	(-)	6	2	3
Fiabilidad				
Disponibilidad de pruebas	(+)	8	3	1
Evaluación anterior	(+)	8	2	2
Reputación del equipo de desarrollo	(+)	8	3	3
Fiabilidad del propósito	(+)	3	3	3
Soporte práctico	(+)	7	3	2
Total puntuación			1.51	1.06

Como resultado de este proceso, en la tabla 4.12 se presentan las ontologías a reutilizar

en cada uno de los grupos planteados para formar la red ontológica UCuenca-Ontology. Cabe mencionar que la ontología VIVO-ISTF será reutilizada en algunos grupos debido a que contiene una gran cantidad de conceptos y propiedades útiles.

Tabla 4.12: Ontologías seleccionadas para ser reutilizadas

Grupo	Ontología a reutilizar
Organización administrativa	<ul style="list-style-type: none">■ Univ-Bench■ VIVO-ISTF
Organización académica	<ul style="list-style-type: none">■ HERO■ VIVO-ISTF
Información personal	<ul style="list-style-type: none">■ FOAF■ VCARD
Espacios físicos	<ul style="list-style-type: none">■ Wgs84 Basic Geo■ VIVO-ISTF
Trabajos de titulación	<ul style="list-style-type: none">■ BIBO

4.2.4. Fusión de recursos ontológicos

Hasta este punto se ha obtenido un conjunto de ontologías de dominio independientes a ser reutilizadas, con la finalidad de fusionarlas para formar una red ontológica se ha realizado un alineamiento manual mediante el cual se han establecido las relaciones que permiten enlazar las distintas ontologías. En la figura 4.2 se muestran las clases y propiedades de cada una de las ontologías seleccionadas que coinciden con los términos planteados en el ORSD y las relaciones establecidas entre ellas.

Como se puede observar, la red ontológica obtenida como resultado contiene parte de los términos definidos en el ORSD, pero no podría responder a todas las preguntas de competencia planteadas ya que carece de conceptos y relaciones fundamentales. Por

ejemplo, la pregunta de competencia 31 "¿Qué docente es responsable de la asignatura X durante un período académico Y?" no puede ser respondida debido a que no existe una relación entre las clases Docente y Curso.

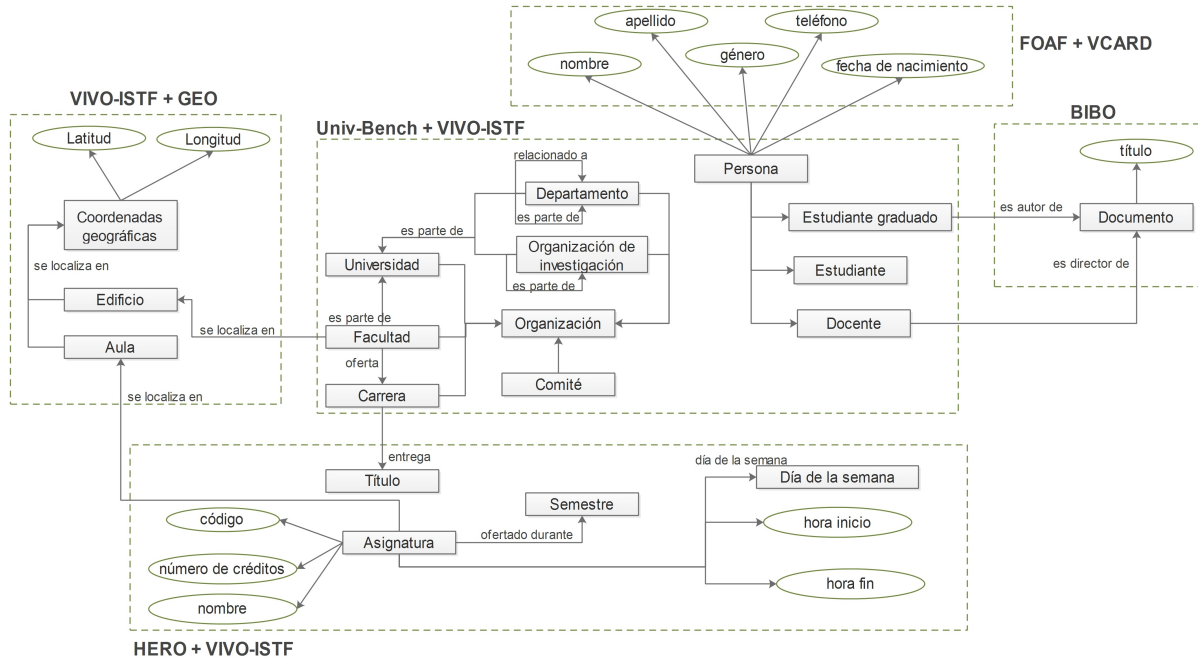


Figura 4.2: Metamodelo de la red ontológica obtenida al fusionar las ontologías seleccionadas

4.2.5. Re-estructuración de recursos ontológicos

Las ontologías seleccionadas no pueden ser reutilizadas tal cual, ya que no contemplan conceptos relevantes establecidos en el ORSD e incluyen otros que son considerados irrelevantes para el dominio a modelar. Con la finalidad de obtener una red ontológica capaz de responder a todas las preguntas de competencia planteadas, se requiere realizar un proceso de re-estructuración de recursos ontológicos.

- **Poda:** En ciertos casos fue necesario eliminar conceptos considerados innecesarios para la red ontológica. Por ejemplo, en el caso de la ontología Univ-Bench se eliminó toda la jerarquía de conceptos relacionados con publicaciones, ya que en el alcance del presente proyecto esto no se considera.
- **Enriquecimiento:** Esta actividad busca incluir conceptos y relaciones que permitan definir los términos del ORSD que no existen en las ontologías seleccionadas. Para esto, y teniendo en consideración el principio de reutilización, se realizó una

búsqueda de términos adecuados para ser reutilizados, únicamente en aquellos casos en los que un término no existe fue creado haciendo uso de la URI establecida para UCuenca-Ontology.

Una vez concluido este proceso se obtuvo como resultado una red ontológica capaz de responder a todas las preguntas definidas en el ORSD, en la figura 4.3 se presenta esta red de forma general, de la cual es importante resumir lo siguiente:

- Se han establecido relaciones que permiten enlazar las partes de la red dedicadas a describir cada uno de los grupos, fusionando de esta manera las diversas ontologías reutilizadas. Se pueden mencionar los siguientes ejemplos:
 - El módulo de localización se enlaza con los módulos correspondientes a la organización administrativa y académica mediante la relación *core:locatedIn*, que permite establecer, en el primer caso, la ubicación de una Facultad dentro del campus universitario y en el segundo caso permite indicar el lugar en el cual se dicta una determinada asignatura de una oferta académica.
 - El módulo correspondiente a la organización administrativa se relaciona con los trabajos de titulación por medio de la relaciones *dbpedia:author* y *bibo:director*, que hacen posible identificar al autor (o autores) y director de un trabajo de titulación específico.
- En una primera etapa de enriquecimiento se reutilizan términos provenientes de diferentes vocabularios, así por ejemplo fueron incluidos los siguientes términos:
 - El concepto *HighSchool* que permite representar el colegio en del cual se graduó como bachiller un estudiante, proviene del vocabulario *Schema*.
 - El concepto *Teacher* que identifica aquellas personas que son docentes dentro de la universidad, proviene de la ontología *TEACH*.
- Posteriormente aquellos términos, específicos para el dominio en cuestión, que no se encontraban definidos en vocabularios existentes fueron creados haciendo uso de la URI establecida para el presente proyecto, tal el es el caso de *uco:AcademicOffer* que representa una oferta académica, *uco:AcademicCareer* que representa una carrera dentro de una Facultad, etc.

El diagrama que incluye todas las clases y propiedades que contiene la red ontológica puede ser observado en el anexo C. Cabe destacar que cada uno de los recursos utilizados incluye el prefijo que indica la ontología a la cual pertenece, las URIs correspondientes a estos prefijos pueden ser encontradas en el Anexo B.

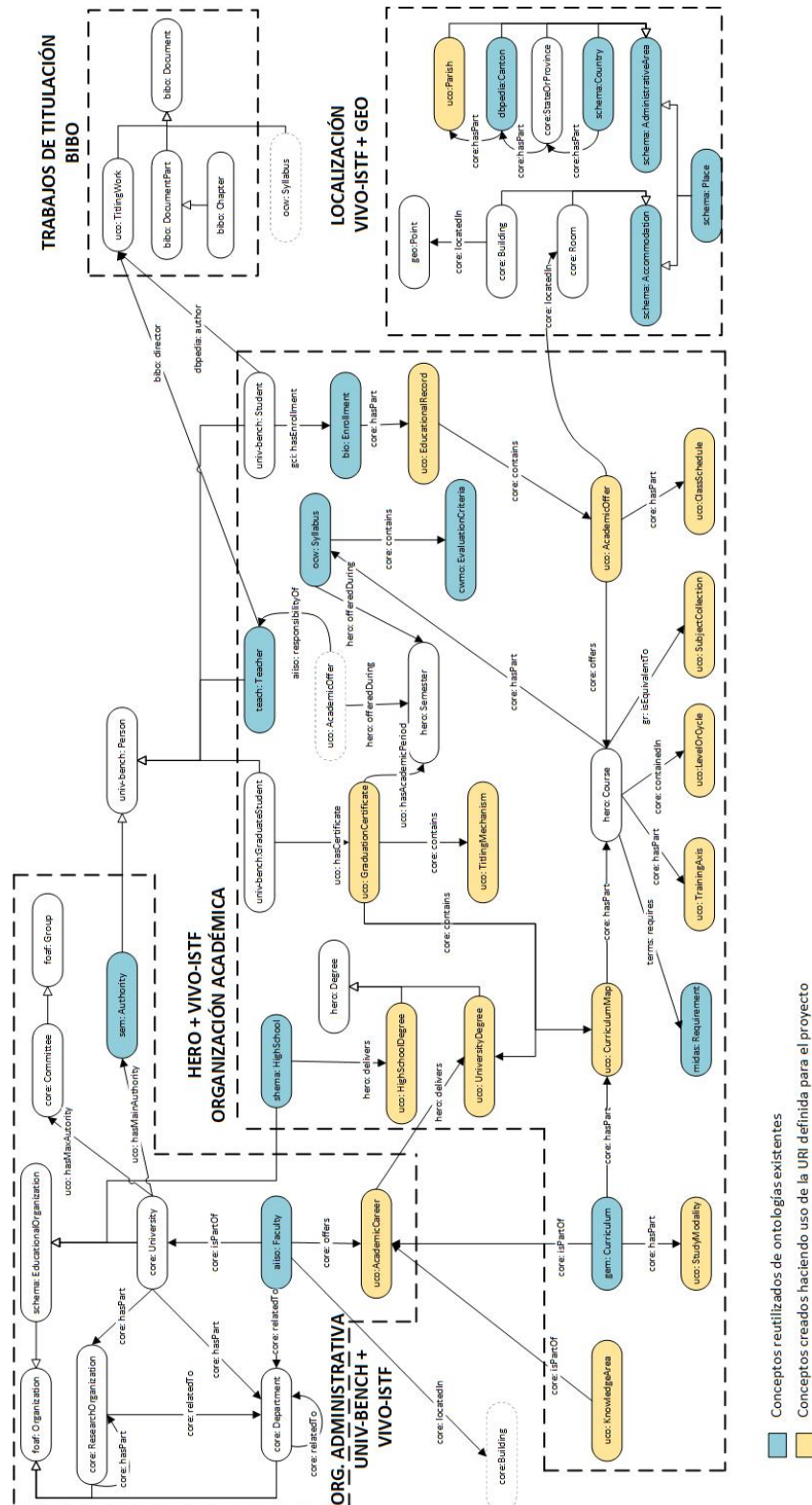


Figura 4.3: Diagrama general de la red ontológica UCuenca-Ontology

4.2.6. Implementación

Considerando los vocabularios seleccionados y las nuevas clases y propiedades incluidas con la finalidad de enriquecer la red ontológica, se procede a modelar e implementar la red ontológica, haciendo uso de la herramienta Protegé v5.3.

Inicialmente se requiere definir la URI a utilizar para el desarrollo de la red ontológica, definir los prefijos a utilizar para cada una de las ontologías e importar aquellas de las cuales se reutilizará un número considerable de recursos, para ello, previamente cada una de dichas ontologías fue podada de forma individual con la finalidad de importar únicamente los términos relevantes para la red ontológica a desarrollar. La figura 4.4 presenta la URI definida, anotaciones básicas agregadas y parte de los vocabularios importados.

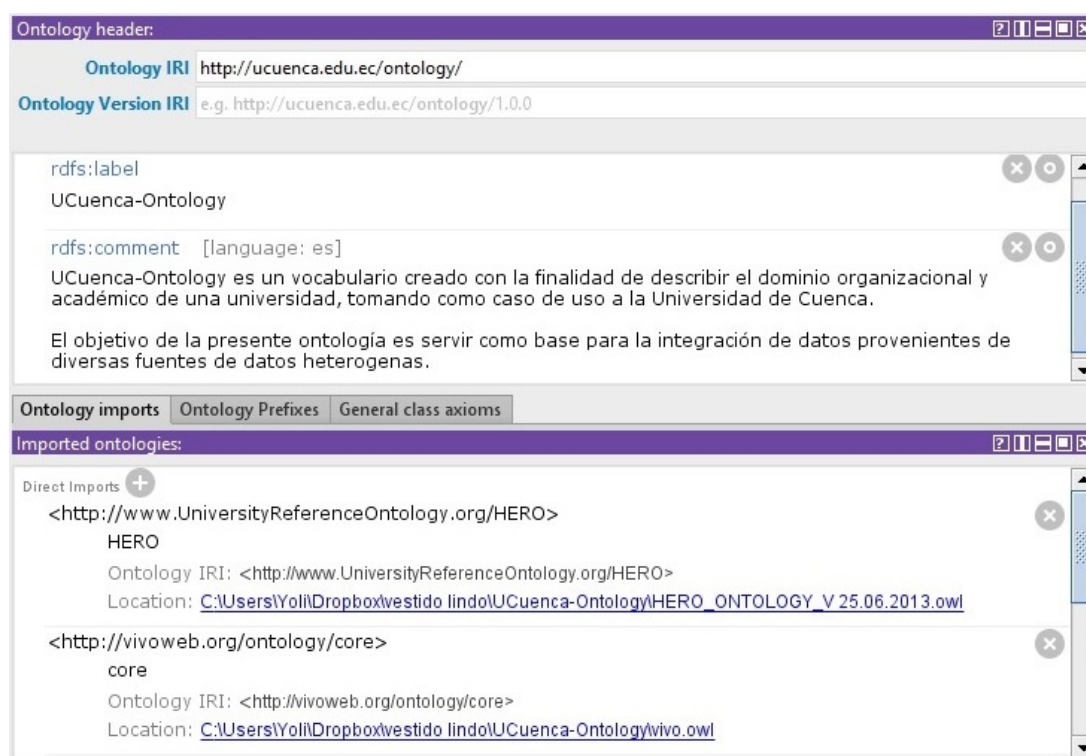


Figura 4.4: Anotaciones y parte de los vocabularios importados a la red ontológica

Luego de definir las ontologías a importar, los conceptos y propiedades que contienen pasan a formar parte del modelo en desarrollo, sin embargo, éstos no incluyen la totalidad de términos que contiene la red ontológica (anexo C), por lo tanto, es necesario crear los términos incluidos como resultado del proceso de enriquecimiento, además establecer

jerarquías en los casos en los que se requiera. En la figura 4.5 se presenta parte de las clases modeladas, el detalle de cada una de las clases, propiedades de objetos y de datos pueden observarse en los anexos D, E y F respectivamente.

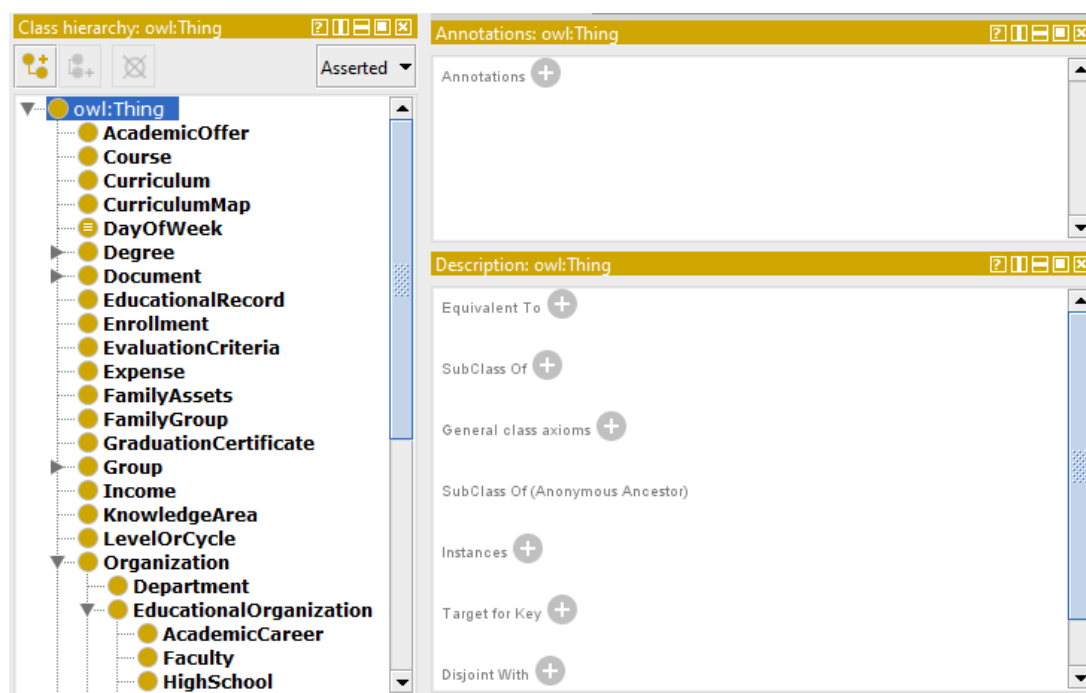


Figura 4.5: Modelado de clases de la red ontológica

Una vez terminado el proceso de creación es necesario establecer las relaciones "Clase - Clase" haciendo uso de las propiedades de objetos y las relaciones "Clase - Dato" mediante las propiedades de datos con la finalidad de formar la red ontológica. En la figura 4.6 se presenta, las relaciones establecidas para la clase Student.

Como último paso, y con el objetivo de facilitar la comprensión de cada uno de los conceptos definidos en la red ontológica, se agregaron anotaciones como definiciones, comentarios y/o etiquetas en idioma español. En la figura 4.7 se muestra las anotaciones incluidas en la clase "AcademicOffer"; adicionalmente en el Anexo G se describen, en forma de tablas, las anotaciones establecidas para las diferentes clases.

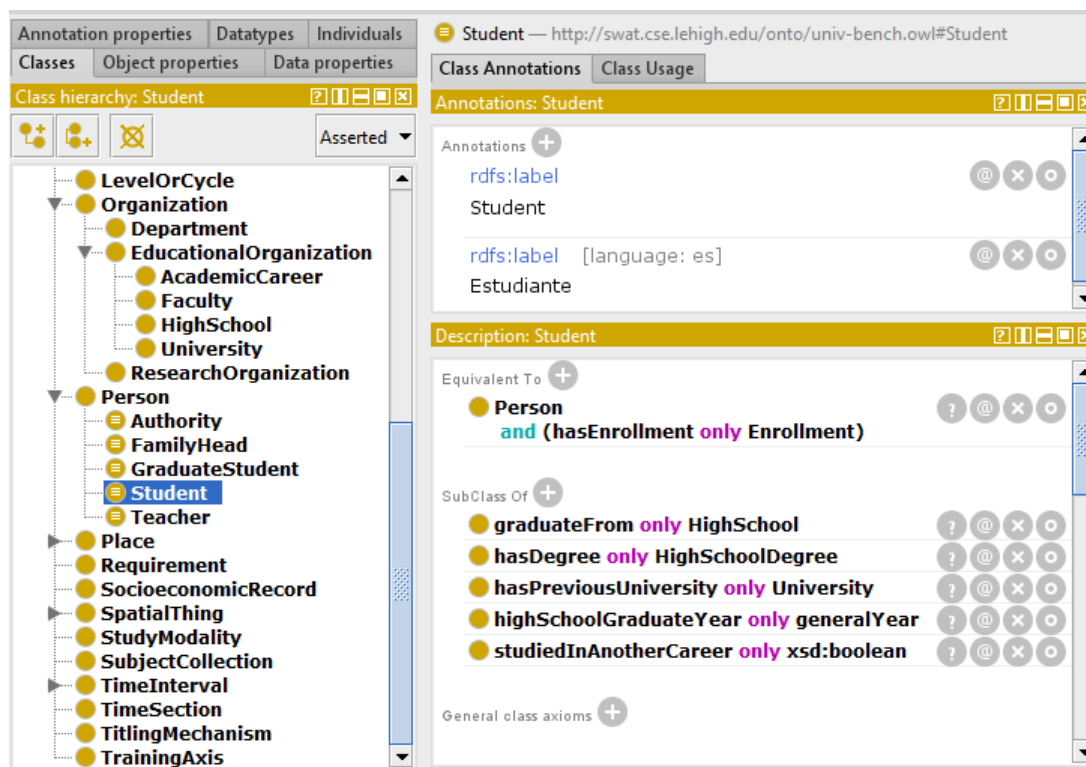


Figura 4.6: Relaciones establecidas para la clase Student

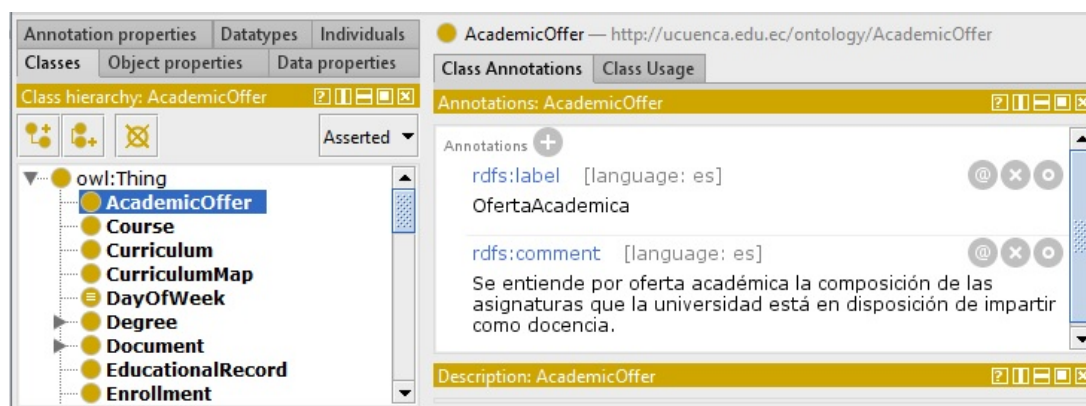


Figura 4.7: Anotaciones incluidas en el concepto "AcademicOffer"

4.3. Generación RDF

Esta etapa tiene como objetivo generar un archivo RDF mediante la ejecución de un proceso semiautomático, el flujo de trabajo se define en el proceso ETL¹³ presentado en la figura 4.8, a continuación se describe cada uno de dichos procesos:

Nota: Con la finalidad de proveer soporte a trabajos similares en el Anexo I se presenta los problemas y las soluciones a los que se enfrentó en esta etapa del proyecto.

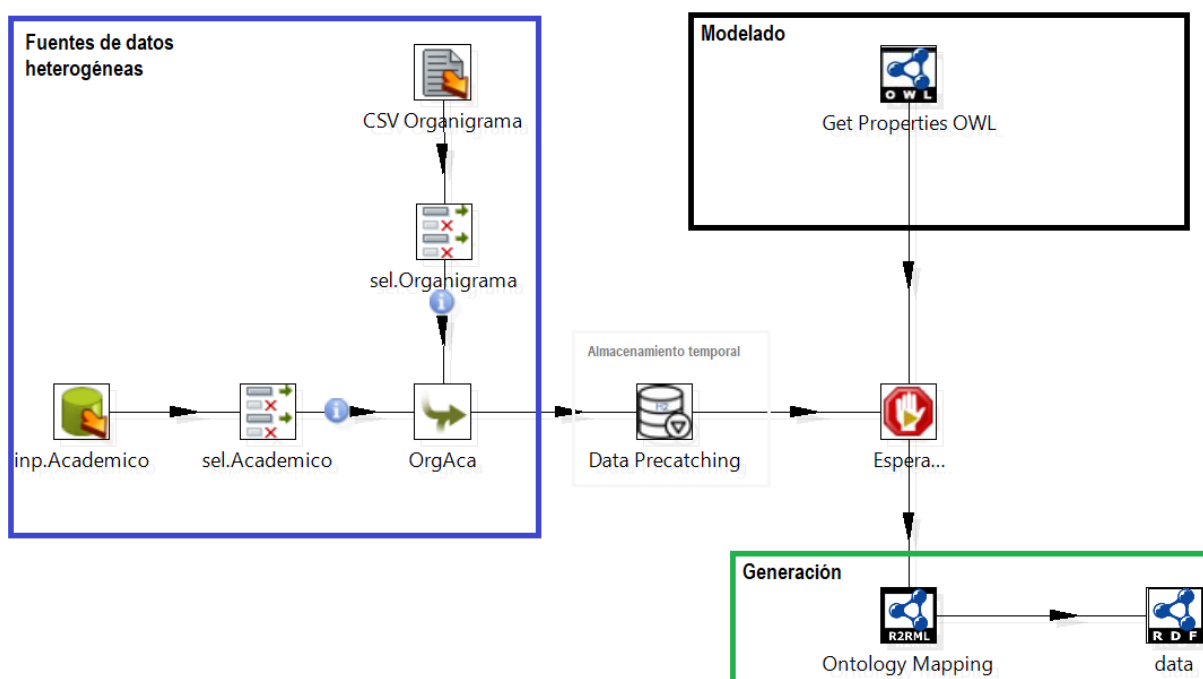


Figura 4.8: Proceso ETL para la generación RDF

Inp.Academico: En la figura 4.9 se presenta el paso que permite recuperar los registros de los esquemas de las bases de datos de Oracle a partir de una consulta SQL (ver Anexo H), dicha consulta retorna los registros en forma de tripletas para facilitar el mapeo en las etapas posteriores.

¹³ETL: Extracción, Transformación y Carga

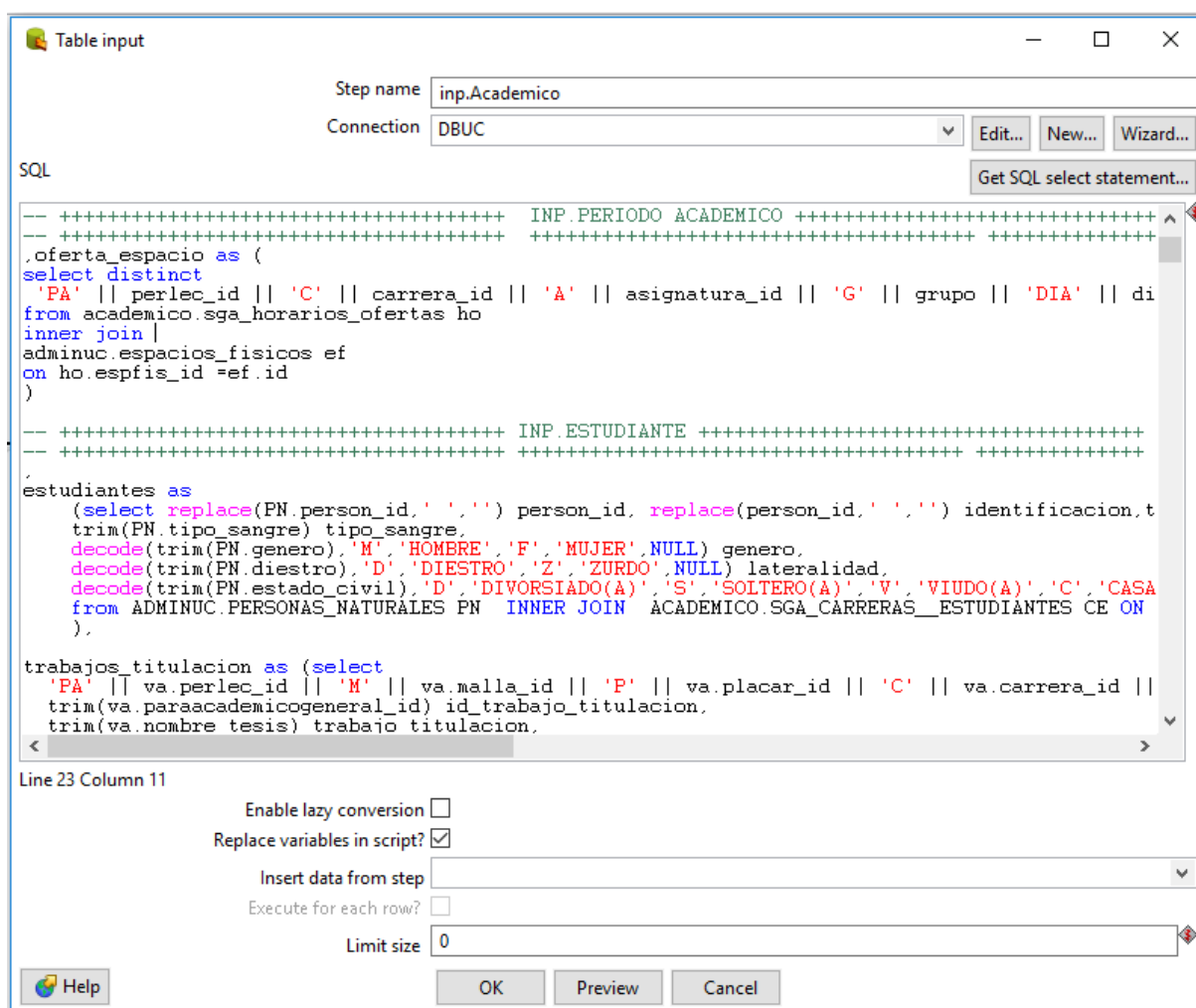
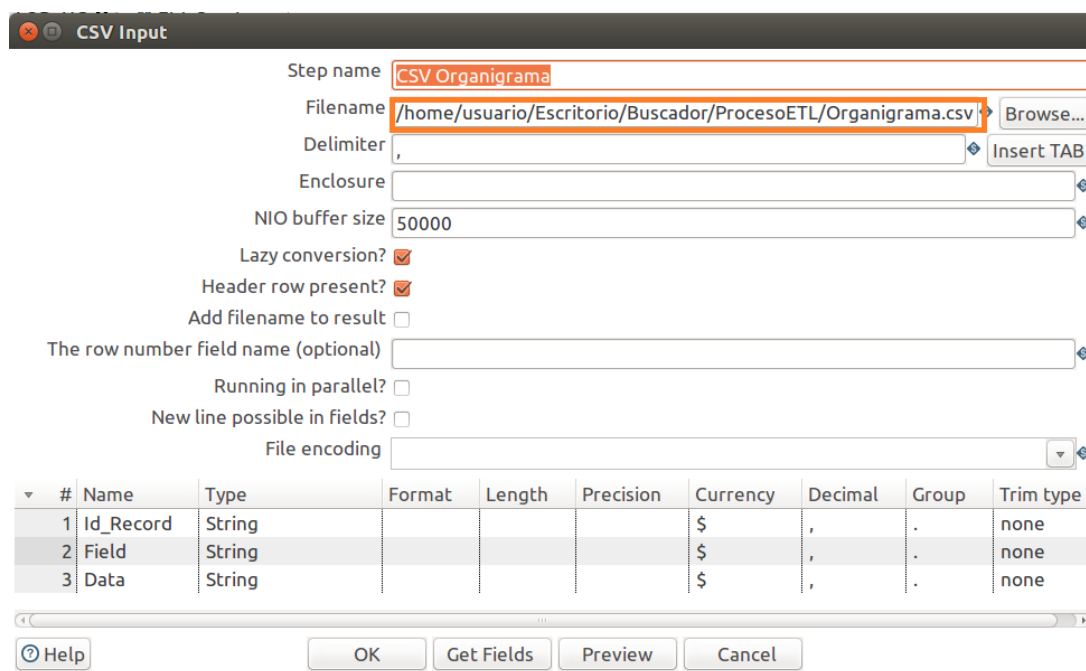


Figura 4.9: Definición del paso para extraer información de las bases de datos Oracle

CSV Organigrama: Realiza la lectura de datos almacenados en un archivo CSV¹⁴ (figura 4.10) que contiene la información necesaria sobre las dependencias que conforman la organización administrativa de la Universidad de Cuenca y sus relaciones en forma de tripleteas.

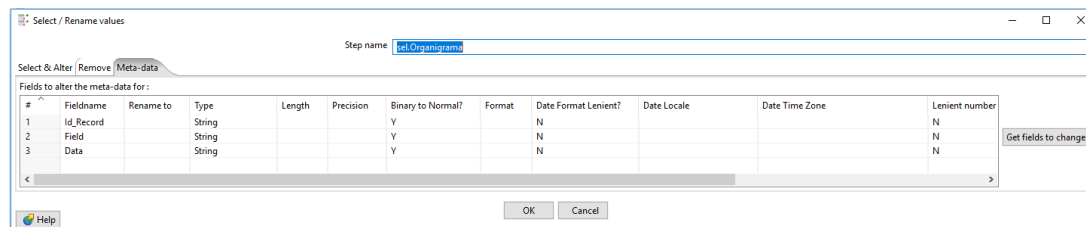
¹⁴CSV (Comma-Separated Values): Formato de documento donde los datos se separan por comas.



#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Id_Record	String				\$,	.	none
2	Field	String				\$,	.	none
3	Data	String				\$,	.	none

Figura 4.10: Lectura de datos de una fuente de extensión .csv

Sel.Organigrama: Se encarga de cambiar el tipo a los datos que provienen del componente CSV Organigrama (figura 4.11).



#	Fieldname	Rename to	Type	Length	Precision	Binary to Normal?	Format	Date Format Lenient?	Date Locale	Date Time Zone	Lenient number
1	Id_Record		String			Y		N			N
2	Field		String			Y		N			N
3	Data		String			Y		N			N

Figura 4.11: Casting de datos para homogeneizar la información

OrgAca: Permite unir conjuntos de datos provenientes de fuentes heterogéneas, en este caso de la base de datos y el archivo CSV. Ver figura 4.12.

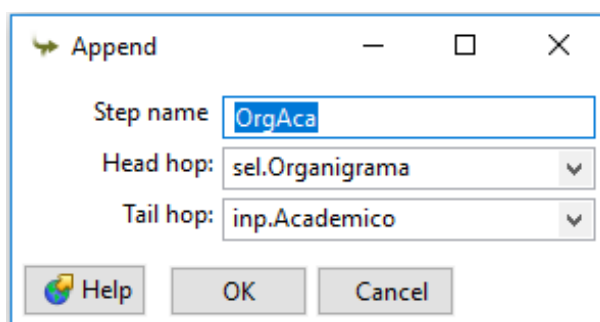


Figura 4.12: Unión de datos provenientes de fuentes heterogéneas

Data Precatching: Este componente, presentado en la figura 4.13, almacena temporalmente la información proveniente de las fuentes heterogéneas con el objetivo de utilizar el estándar de mapeo R2RML¹⁵.

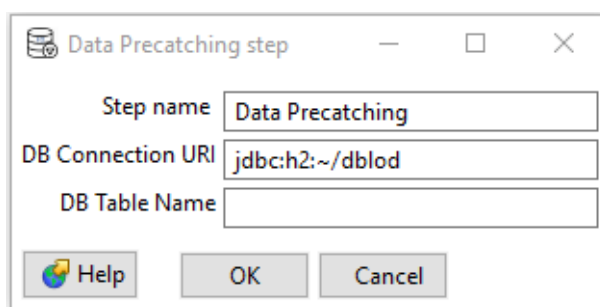


Figura 4.13: Conexión a bases de datos embebida para el almacenamiento temporal de datos

Get Properties OWL: Lee un archivo OWL generado en protégé, el archivo a especificar contiene las clases y propiedades a las cuales serán mapeados los datos provenientes de las fuentes heterogéneas. En el componente de la figura 4.14 se especifica la ruta del archivo ontológico (OWL).

¹⁵R2RML: Lenguaje que permite expresar mapeos personalizados de bases de datos relacionales a conjuntos de datos RDF[44]

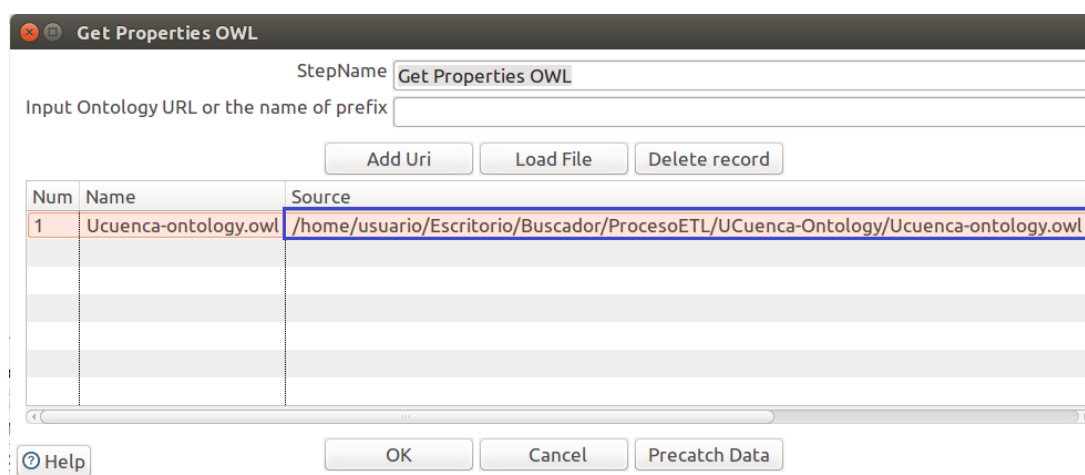


Figura 4.14: Lectura del modelo ontológico desde archivo con extensión OWL

Espera...: Pausa temporalmente el flujo hasta que los pasos establecidos en el componente hayan finalizado. En la figura 4.15 se ha establecido que el flujo se detenga hasta que los pasos Data Precatching y Get Properties OWL hayan concluido.

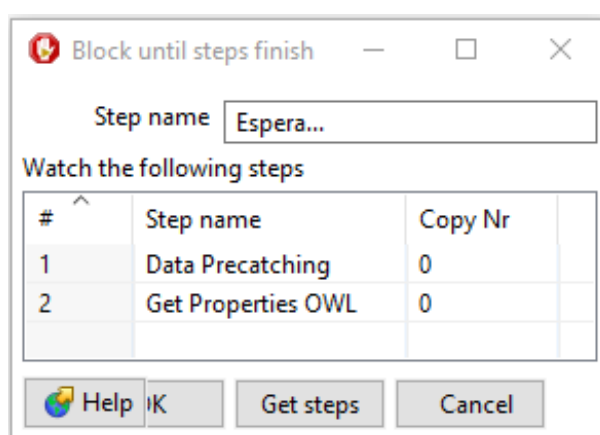


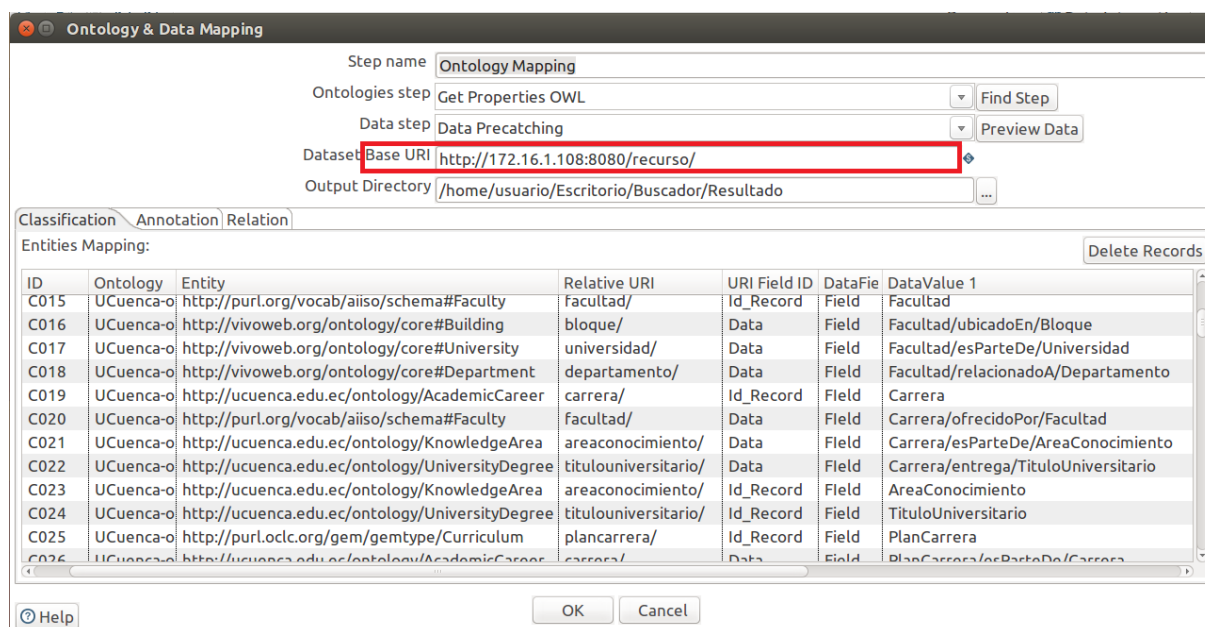
Figura 4.15: Bloqueo de flujo debido a dependencias entre pasos en PDI

Ontology Mapping: En este paso se establecen las reglas de mapeo, es decir, se especifica los registros que se transformarán en clases, propiedades de objeto o propiedades de datos. La salida obtenida es un archivo R2RML. Cabe mencionar que este es uno de los pasos más importantes de la transformación, debido a que en este punto se hace uso de varias salidas de las etapas anteriores como: URI base diseñada en la etapa de especificación, red ontológica definida en la etapa de modelado y los datos depurados provenientes de las fuentes heterogéneas.

Es importante aclarar que, a pesar de que la URI base establecida en la sección 4.1.3 para las instancias fue `http://ucuenca.edu.ec/resource/{resourcetype}/{resourcename}`, tuvo que ser modificada a `http://172.16.1.108:8080/resource/{resourcetype}/{resourcename}` para el caso de prueba, debido a que en las etapas posteriores las peticiones de información semántica se realizarán al servidor en el cual se encuentran alojadas las instancias generadas en esta etapa.

A continuación, se presenta una breve descripción acerca de las reglas de mapeo entre la fuente de datos y los términos del modelo ontológico.

En la figura 4.16 se presenta las reglas de mapeo establecidas entre los datos y clases. El propósito de este paso es establecer los registros que se convertirán en un recurso del tipo clase.



Step name:

Ontologies step:

Data step:

Dataset Base URI:

Output Directory:

Classification Annotation Relation

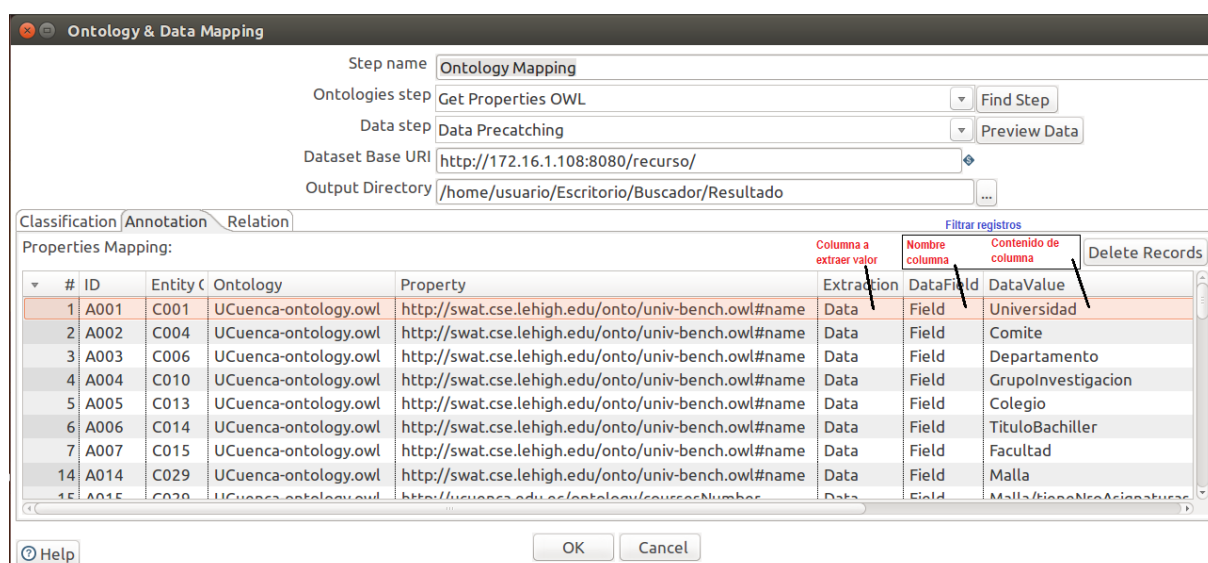
Entities Mapping:

ID	Ontology	Entity	Relative URI	URI Field ID	DataFile	DataValue 1
C015	UCuenca-o	http://purl.org/vocab/aiiso/schema#Faculty	Facultad/	Id_Record	Field	Facultad
C016	UCuenca-o	http://vivoweb.org/ontology/core#Building	bloque/	Data	Field	Facultad/ubicadoEn/Bloque
C017	UCuenca-o	http://vivoweb.org/ontology/core#University	universidad/	Data	Field	Facultad/esParteDe/Universidad
C018	UCuenca-o	http://vivoweb.org/ontology/core#Department	departamento/	Data	Field	Facultad/relacionadoA/Departamento
C019	UCuenca-o	http://ucuenca.edu.ec/ontology/AcademicCareer	carrera/	Id_Record	Field	Carrera
C020	UCuenca-o	http://purl.org/vocab/aiiso/schema#Faculty	facultad/	Data	Field	Carrera/ofrecidoPor/Facultad
C021	UCuenca-o	http://ucuenca.edu.ec/ontology/KnowledgeArea	areaconocimiento/	Data	Field	Carrera/esParteDe/AreaConocimiento
C022	UCuenca-o	http://ucuenca.edu.ec/ontology/UniversityDegree	titulouniversitario/	Data	Field	Carrera/entrega/TituloUniversitario
C023	UCuenca-o	http://ucuenca.edu.ec/ontology/KnowledgeArea	areaconocimiento/	Id_Record	Field	AreaConocimiento
C024	UCuenca-o	http://ucuenca.edu.ec/ontology/UniversityDegree	titulouniversitario/	Id_Record	Field	TituloUniversitario
C025	UCuenca-o	http://purl.oclc.org/gem/gemtype/Curriculum	plancarrera/	Id_Record	Field	PlanCarrera
C026	UCuenca-o	http://ucuenca.edu.ec/ontology/AcademicCareer	carrera/	Data	Field	PlanCarrera/esParteDe/Carrera

OK Cancel

Figura 4.16: Reglas de mapeo entre datos y clases

En la figura 4.17 se presenta las reglas de mapeo establecidas entre los datos y propiedades de datos. El propósito de este paso es definir los registros que se transformarán en propiedades de datos de un determinado recurso.



Step name:

Ontologies step:

Data step:

Dataset Base URI:

Output Directory:

Classification / Annotation / Relation

Properties Mapping:

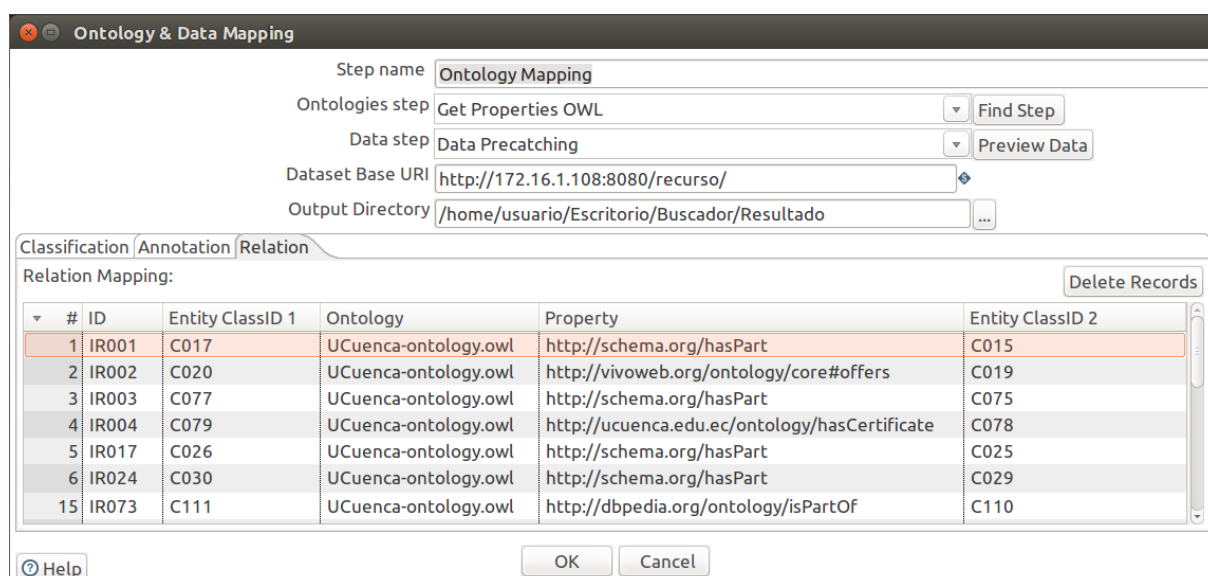
#	ID	Entity	Ontology	Property	Extration	DataField	DataValue
1	A001	C001	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Universidad
2	A002	C004	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Comite
3	A003	C006	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Departamento
4	A004	C010	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	GrupoInvestigacion
5	A005	C013	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Colegio
6	A006	C014	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	TituloBachiller
7	A007	C015	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Facultad
14	A014	C029	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Malla
15	A015	C020	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#name	Data	Field	Malla/lineaNoAsignatura

Annotations: Columna a extraer valor (points to Extration), Nombre columna (points to DataField), Contenido de columna (points to DataValue).

Buttons: Help, OK, Cancel

Figura 4.17: Reglas de mapeo entre datos y propiedades de datos

En la figura 4.18 se presenta las reglas de mapeo establecidas entre los datos y las propiedades de objetos. El propósito de este paso es relacionar entidades a través del uso de las propiedades de objeto definidas en el modelo ontológico.



Step name:

Ontologies step:

Data step:

Dataset Base URI:

Output Directory:

Classification / Annotation / Relation

Relation Mapping:

#	ID	Entity ClassID 1	Ontology	Property	Entity ClassID 2
1	IR001	C017	UCuenca-ontology.owl	http://schema.org/hasPart	C015
2	IR002	C020	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#offers	C019
3	IR003	C077	UCuenca-ontology.owl	http://schema.org/hasPart	C075
4	IR004	C079	UCuenca-ontology.owl	http://ucuenca.edu.ec/ontology/hasCertificate	C078
5	IR017	C026	UCuenca-ontology.owl	http://schema.org/hasPart	C025
6	IR024	C030	UCuenca-ontology.owl	http://schema.org/hasPart	C029
15	IR073	C111	UCuenca-ontology.owl	http://dbpedia.org/ontology/isPartOf	C110

Buttons: Help, OK, Cancel

Figura 4.18: Reglas de mapeo entre datos y propiedades de objetos

data: En este paso se generan las tripletas RDF tomando como entrada el archivo R2RML. El componente de la figura 4.19 presenta los parámetros de configuración nece-

sarios para la generación del RDF, la mayor parte de estos parámetros son definidos a través de la opción «Retrieve DB Connection from input Step», por tanto los parámetros a definir manualmente son: ruta donde se almacenará el RDF, y el formato de presentación de tripletas.

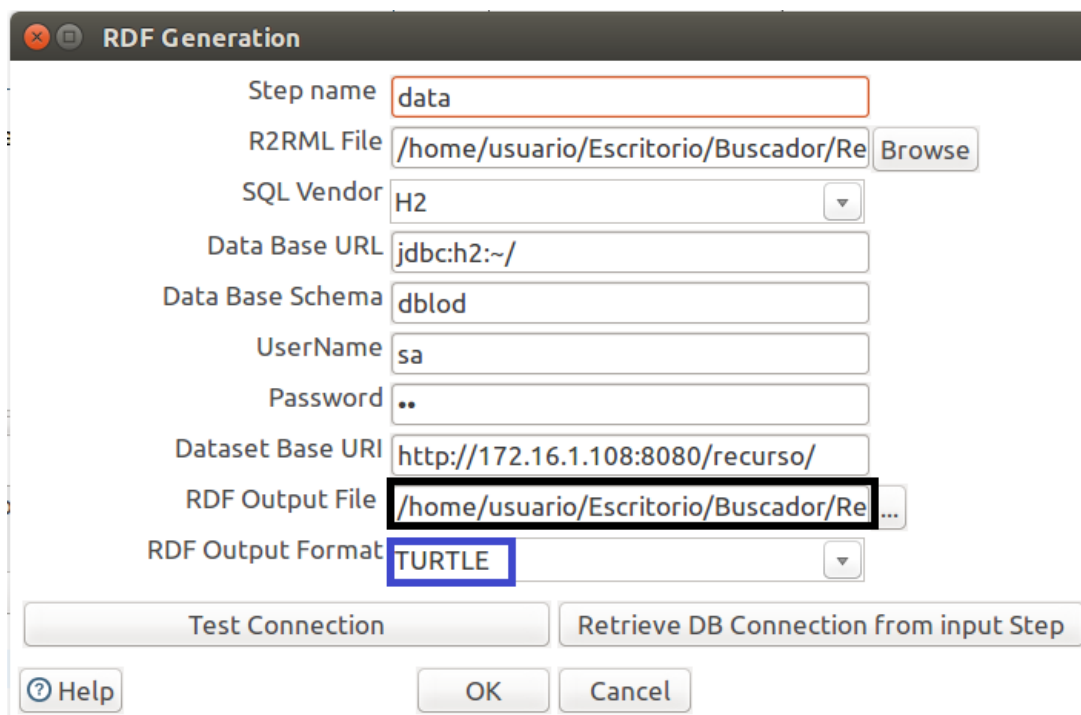


Figura 4.19: Configuración del componente RDF Generation

4.4. Publicación

Haciendo uso del plugin Fuseki Loader se ha desplegado el triplestore Fuseki que permite almacenar el recurso RDF generado en el paso anterior (ver figura 4.19), además ofrece un servicio de acceso SPARQL Endpoint, tal como se muestra en la figura 4.22.

El despliegue del servicio antes mencionado tiene el objetivo de permitir la ejecución de búsquedas de información a través de sentencias SPARQL mediante solicitudes HTTP, lo cual será utilizado posteriormente por el buscador semántico. Además requiere la configuración de los archivos `scr2.sh` y `UDC.ttl`, localizados en el proyecto FUSEKI2 (ver figura 4.20), debido a que:

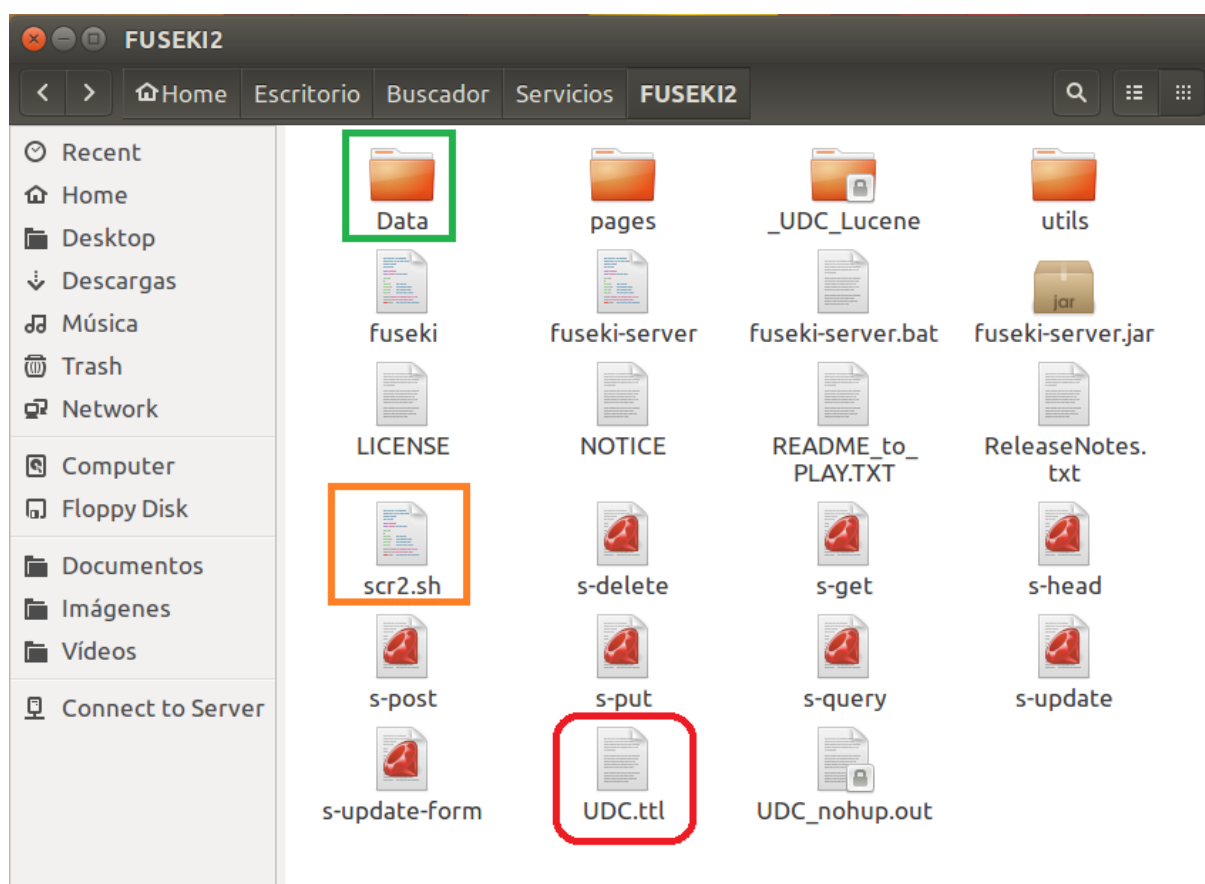
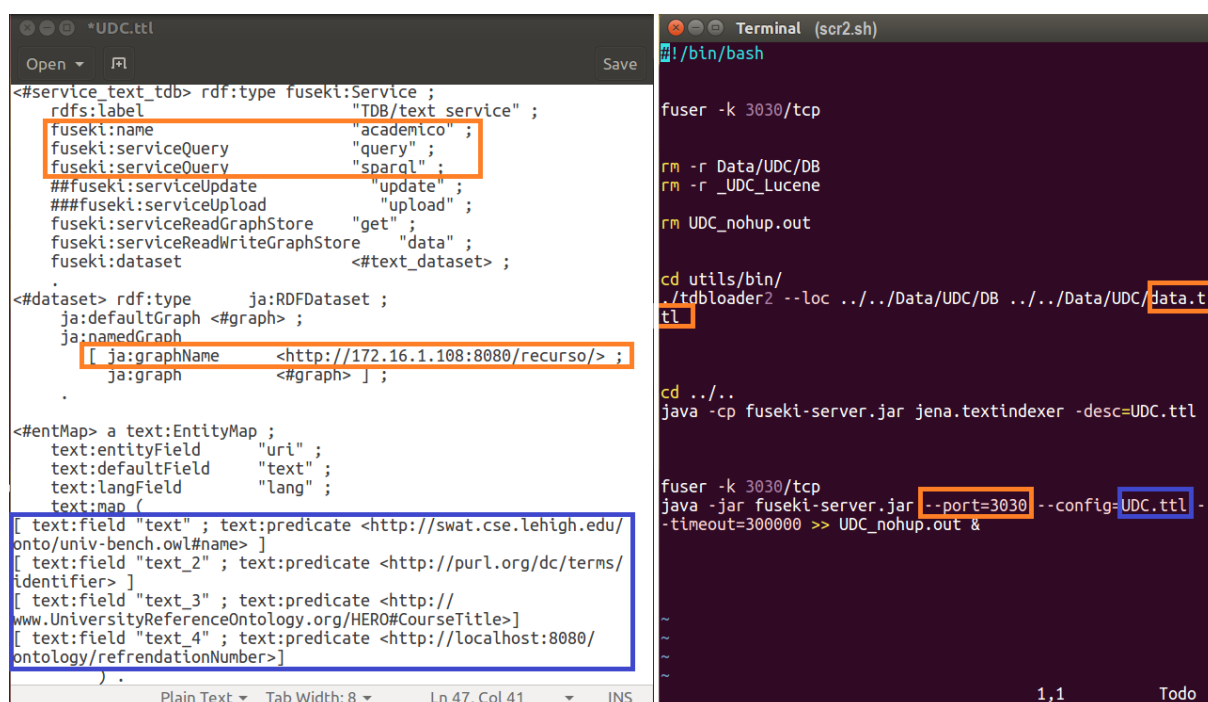


Figura 4.20: Proyecto Fuseki

- En el archivo UDC.ttl se especifica el nombre del servicio con el cual se realizará determinadas consultas SPARQL vía HTTP, además se definirán las propiedades que se indexarán para obtener resultados eficaces. Tal configuración se presenta en la ventana izquierda de la figura 4.21.
- El archivo scr2.sh contiene un script para desplegar el servicio Fuseki Endpoint, por tanto los parametros relevantes a especificar en dicho archivo son: el puerto en el cual se ejecutará el servicio; ruta del archivo de configuración UDC.ttl; y el archivo RDF que se generó en la etapa de Generación RDF. Tal configuración se presenta en la ventana derecha de la figura 4.21.



The image shows a text editor window on the left and a terminal window on the right. The text editor contains an RDF configuration file named `UDC.ttl`. The configuration defines a service named `text_tdb` of type `fuseki:Service` with the label `TDB/text service`. It specifies the service name as `academico`, the query language as `query`, and the SPARQL endpoint as `sparql`. It also defines update, upload, and get operations. The dataset is named `text_dataset`. The graph is named `graph` with the URI `http://172.16.1.108:8080/recurso/`. The entity map is defined with fields `text`, `text_2`, `text_3`, and `text_4`, each with a specific predicate. The terminal window shows the execution of the `scr2.sh` script, which sets up the environment, removes old data, and starts the Fuseki server on port 3030 using the `UDC.ttl` configuration file.

```
<#service_text_tdb> rdf:type fuseki:Service ;
  rdfs:label "TDB/text service" ;
  fuseki:name "academico" ;
  fuseki:serviceQuery "query" ;
  fuseki:serviceQuery "sparql" ;
  ##fuseki:serviceUpdate "update" ;
  ###fuseki:serviceUpload "upload" ;
  fuseki:serviceReadGraphStore "get" ;
  fuseki:serviceReadWriteGraphStore "data" ;
  fuseki:dataset <#text_dataset> ;
.

<#dataset> rdf:type ja:RDFDataset ;
  ja:defaultGraph <#graph> ;
  ja:namedGraph
    [ ja:graphName <http://172.16.1.108:8080/recurso/> ;
      ja:graph <#graph> ] ;
.

<#entMap> a text:EntityMap ;
  text:entityField "uri" ;
  text:defaultField "text" ;
  text:langField "lang" ;
  text:map (
    [ text:field "text" ; text:predicate <http://swat.cse.lehigh.edu/
      onto/univ-bench.owl#name> ]
    [ text:field "text_2" ; text:predicate <http://purl.org/dc/terms/
      identifier> ]
    [ text:field "text_3" ; text:predicate <http://
      www.UniversityReferenceOntology.org/HERO#CourseTitle> ]
    [ text:field "text_4" ; text:predicate <http://localhost:8080/
      ontology/refrendationNumber> ]
  ) .
```

```
Terminal (scr2.sh)
~/bin/bash
fuser -k 3030/tcp
rm -r Data/UDC/DB
rm -r _UDC_Lucene
rm UDC_nohup.out

cd utils/bin/
./tdbloader2 --loc ../../Data/UDC/DB ../../Data/UDC/data.t
tl

cd ../../
java -cp fuseki-server.jar jena.textindexer -desc=UDC.ttl

fuser -k 3030/tcp
java -jar fuseki-server.jar --port=3030 --config=UDC.ttl -
-timeout=300000 >> UDC_nohup.out &
```

Figura 4.21: Presentación de contenido - Servicio Elda

Finalmente el servicio se despliega ejecutando el script por medio del comando «`sudo sh scr2.sh`». El servicio estará disponible en `http://172.16.1.108:3030`, y su apariencia será similar al que se presenta en la figura 4.22.

Además, en la figura 4.22 se presenta a modo de ejemplo una consulta SPARQL para recuperar los nombres de las carreras que oferta la Universidad de Cuenca.

Fuseki Query

Dataset: /academico

SPARQL Query

```
PREFIX uy:<http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX hero:<http://www.UniversityReferenceOntology.org/HERO#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX core:<http://vivoweb.org/ontology/core#>
select ?nombre
where{?carrera rdf:type core:School.
?carrera uy:name ?nombre
}
ORDER BY ASC(?nombre)
```

Output: Text

If XML output, add XSLT style sheet (blank for none):

☐ Force the accept header to text/plain regardless.

Get Results

SPARQL Update

Perform update

"ADMINISTRACION DE EMPRESAS"

"ARQUITECTURA"

"ARTES APLICADAS"

"ARTES ESCENICAS"

"ARTES MUSICALES"

"ARTES PLASTICAS"

"ARTES VISUALES"

"BIOESTADISTICA"

"BIOQUIMICA Y FARMACIA"

"BIOQUIMICA Y FARMACIA-REDISEÑO"

"CANTO"

"CIENCIAS DE LA COMUNICACIÓN SOCIAL EN COMUNICACIÓN ORGANIZACIONAL Y RELACIONES"

"CIENCIAS DE LA COMUNICACIÓN SOCIAL EN PERIODISMO Y COMUNICACIÓN DIGITAL"

"CIENCIAS DE LA COMUNICACIÓN SOCIAL MENCIÓN EN PERIODISMO"

"CIENCIAS DE LA COMUNICACIÓN SOCIAL MENCIÓN EN PUBLICIDAD Y RELACIONES PÚBLICAS"

"CIENCIAS DE LA EDUCACION E INVESTIGACION DE LAS CULTURAS AMAZONICAS"

"CIENCIAS DE LA EDUCACION EN LA ESPECIALIZACION DE CASTELLANO Y LITERATURA"

"CIENCIAS DE LA EDUCACION EN LA ESPECIALIZACION DE CIENCIAS NATURALES"

"CIENCIAS DE LA EDUCACION EN LA ESPECIALIZACION DE CIENCIAS SOCIALES"

"CIENCIAS DE LA EDUCACION EN LA ESPECIALIZACION DE CULTURA FISICA"

"CIENCIAS DE LA EDUCACION EN LA ESPECIALIZACION DE DIBUJO Y PINTURA"

Figura 4.22: Fuseki Endpoint

4.5. Explotación

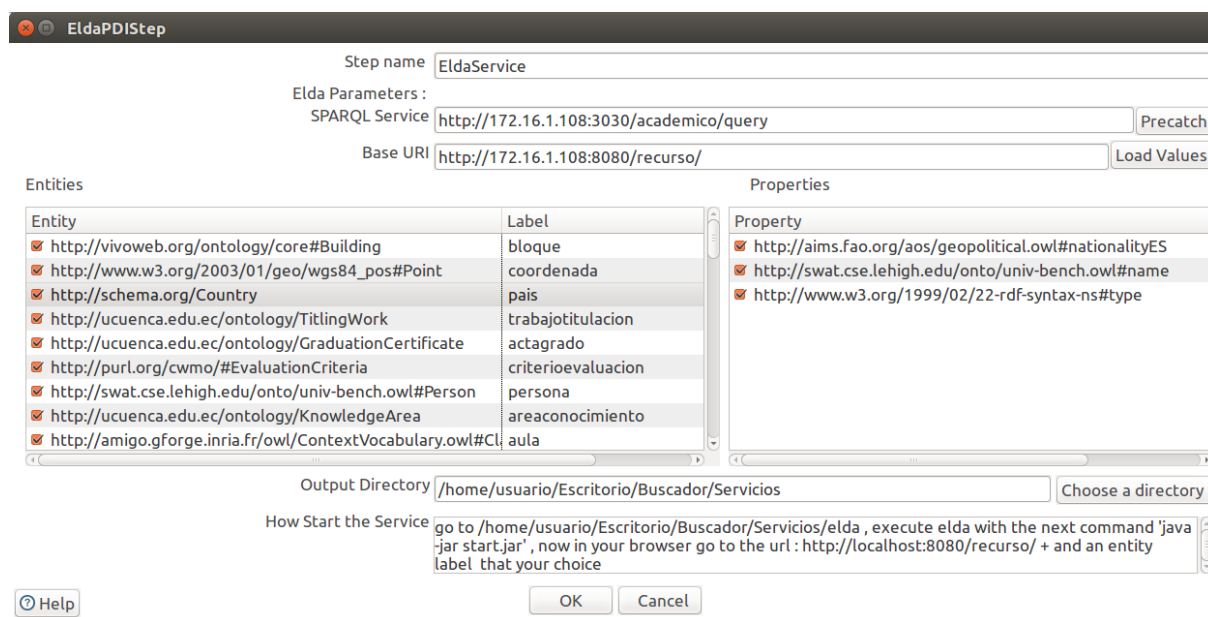
En esta etapa se ha hecho uso del plugin Elda que tiene por objetivo permitir la visualización de contenido (en formato HTML, XML, JSON, RDF/XML y turtle) RDF a través de URLs, cabe mencionar que para su funcionamiento el servicio Fuseki Endpoint debe encontrarse activo debido a que se requiere la ejecución de consultas SPARQL para obtener la información que presentará el servicio Elda.

En la figura 4.23 se presenta la configuración realizada en el plugin Elda de PDI para la generación de un proyecto de este tipo, dichas configuraciones son:

- Especificación del path relativo para entidades.
- Definición de etiquetas adecuadas para las propiedades de datos y propiedades de objeto.
- Especificación de la URI base y nombre del servicio SPARQL de fuseki endpoint.

Yolanda Aucapiña, Carlos Plaza

101



Step name: EldaService

Elda Parameters:

SPARQL Service:

Base URI:

Entities

Entity	Label
<input checked="" type="checkbox"/> http://vivoweb.org/ontology/core#Building	bloque
<input checked="" type="checkbox"/> http://www.w3.org/2003/01/geo/wgs84_pos#Point	coordenada
<input checked="" type="checkbox"/> http://schema.org/Country	pais
<input checked="" type="checkbox"/> http://ucuenca.edu.ec/ontology/TitlingWork	trabajotitulacion
<input checked="" type="checkbox"/> http://ucuenca.edu.ec/ontology/GraduationCertificate	actagrado
<input checked="" type="checkbox"/> http://purl.org/cwmo/#EvaluationCriteria	criterioevaluacion
<input checked="" type="checkbox"/> http://swat.cse.lehigh.edu/onto/univ-bench.owl#Person	persona
<input checked="" type="checkbox"/> http://ucuenca.edu.ec/ontology/KnowledgeArea	areaconocimiento
<input checked="" type="checkbox"/> http://amigo.gforge.inria.fr/owl/ContextVocabulary.owl#Cl	aula

Properties

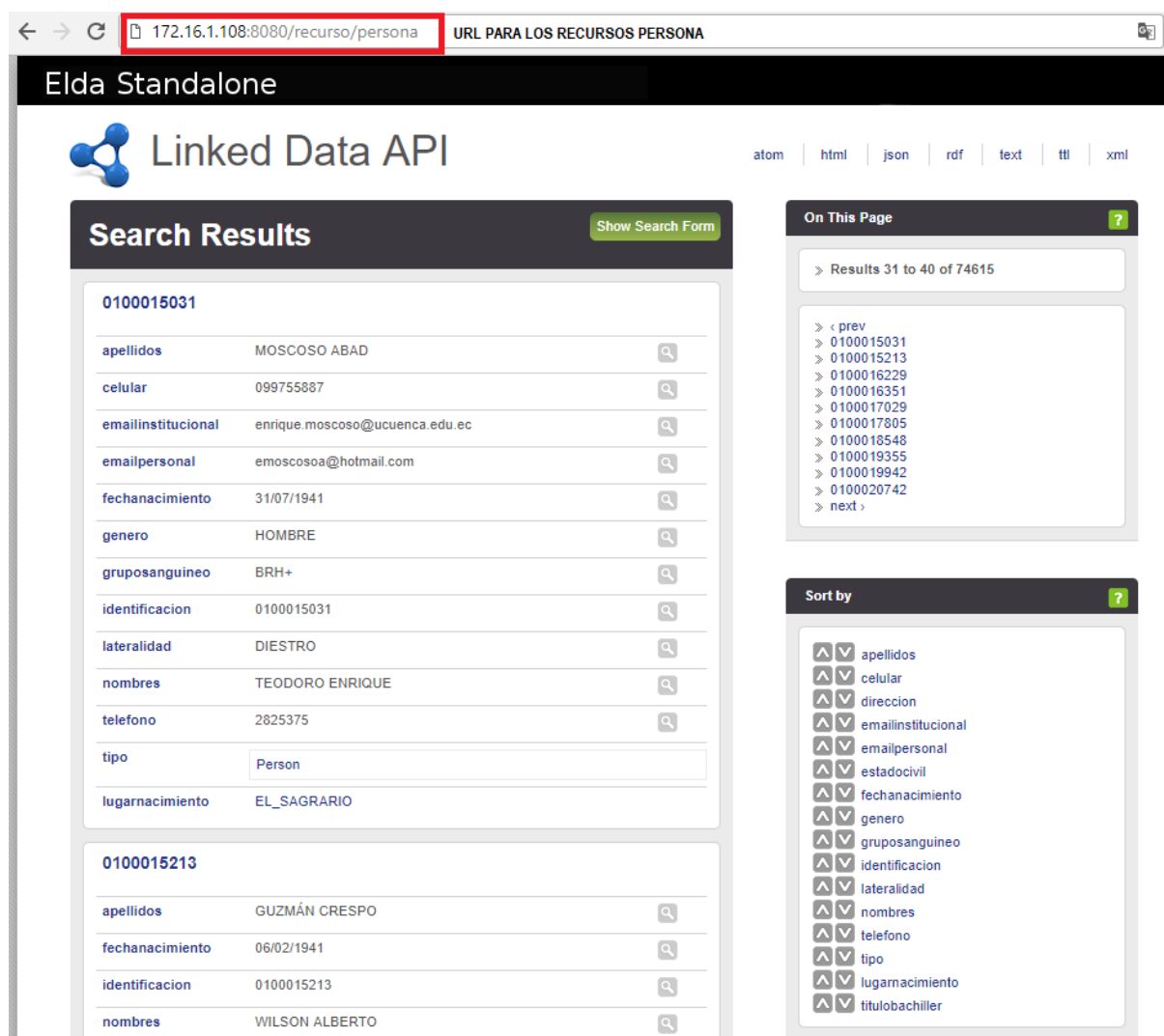
Property
<input checked="" type="checkbox"/> http://aims.fao.org/aos/geopolitical.owl#nationalityES
<input checked="" type="checkbox"/> http://swat.cse.lehigh.edu/onto/univ-bench.owl#name
<input checked="" type="checkbox"/> http://www.w3.org/1999/02/22-rdf-syntax-ns#type

Output Directory:

How Start the Service: go to /home/usuario/Escritorio/Buscador/Servicios/elda , execute elda with the next command 'java -jar start.jar' , now in your browser go to the url : http://localhost:8080/recurso/ + and an entity label that your choice

Figura 4.23: Configuración del plugin para la generación de un proyecto Elda

Una vez que el proyecto ha sido generado situarse en la carpeta raíz y ejecutar el comando «*sudo java -jar start.jar*». EL resultado de este proceso es el despliegue del servicio Elda con lo que se podrá acceder a la información a partir de una URL. A modo de ejemplo en la figura 4.24 se presenta la información acerca de los recursos del tipo *Person* a partir de la URL *http:172.16.1.108:8080/recurso/persona*.



The screenshot displays the Elda Standalone Linked Data API interface. The browser address bar shows the URL `172.16.1.108:8080/recurso/persona`. The page title is "Elda Standalone". The main header features the "Linked Data API" logo and navigation links for `atom`, `html`, `json`, `rdf`, `text`, `ttl`, and `xml`.

The "Search Results" section displays two individual records:

0100015031

apellidos	MOSCOSO ABAD
celular	099755887
emailinstitucional	enrique.moscoso@ucuenca.edu.ec
emailpersonal	emoscosoa@hotmail.com
fechanacimiento	31/07/1941
genero	HOMBRE
gruposanguineo	BRH+
identificacion	0100015031
lateralidad	DIESTRO
nombres	TEODORO ENRIQUE
telefono	2825375
tipo	Person
lugarnacimiento	EL_SAGRARIO

0100015213

apellidos	GUZMÁN CRESPO
fechanacimiento	06/02/1941
identificacion	0100015213
nombres	WILSON ALBERTO

The "On This Page" section shows a list of results from 31 to 40 of 74615, with navigation links for previous and next results.

The "Sort by" section provides a list of attributes for sorting, including `apellidos`, `celular`, `direccion`, `emailinstitucional`, `emailpersonal`, `estadocivil`, `fechanacimiento`, `genero`, `gruposanguineo`, `identificacion`, `lateralidad`, `nombres`, `telefono`, `tipo`, `lugarnacimiento`, and `titulobachiller`.

Figura 4.24: Presentación de contenido - Servicio Elda

4.6. Resultados

En base al proceso llevado a cabo durante cada una de las actividades descritas en las secciones anteriores se puede recalcar los siguientes resultados:

- **Información anotada semánticamente:** Debido al uso de tecnología semántica (ontologías, repositorios, etc.), y reglas de mapeo realizadas en la etapa de generación.
- **Datos integrados:** La información proveniente de fuentes heterogéneas se ha

integrado con un único formato.

- **Modelo navegable e interoperable:** RDF, al ser un modelo de datos estándar en la web semántica, puede ser utilizado por diferentes aplicaciones. En la figura 4.25 se presenta un extracto de tripleas RDF.



Figura 4.25: Extracto RDF - Información Universidad de Cuenca

- **Obtención de nuevo conocimiento:** A diferencia de una base de datos tradicional, una ontología provee la posibilidad de generar nuevo conocimiento a partir de las restricciones e instancias definidas en el modelo. Con la finalidad de demostrar esta característica, se ha cargado las instancias generadas en PDI al modelo ontológico a través de la API Jena ¹⁶, para posteriormente almacenarlas con formato RDF y ejecutarlas en Protegé.

Para llevar a cabo esta demostración se ha limitado la cantidad de datos considerando únicamente una determinada parte de la información disponible de la carrera de Ingeniería de Sistemas, y se ha generado mediante PDI un archivo TTL ¹⁷ con instancias para el modelo ontológico con las siguientes características:

- Las instancias referentes a las personas han sido generadas teniendo en cuenta la entidad *Person*, por lo tanto en el archivo de instancias no es posible diferenciar a los docentes, estudiantes y jefes de familia.

¹⁶Apache Jena: Framework Java que permite construir aplicaciones basadas en ontologías.

¹⁷TTL: Extensión de un archivo turtle para la representación textual de un grafo RDF.

- Se utiliza únicamente una propiedad de objeto para enlazar dos instancias de clases, es decir no se ha generado instancias con propiedades inversas.

Dicho esto, se tiene por objetivo diferenciar a las personas como estudiantes, docentes o jefes de familia. Esto es posible, debido a que en el modelo ontológico las clases correspondientes fueron modeladas como clases definidas, es decir, contienen condiciones de equivalencia; por ejemplo, en la figura 4.26 puede observarse que la clase *Student* es equivalente a una persona que posee una matrícula.

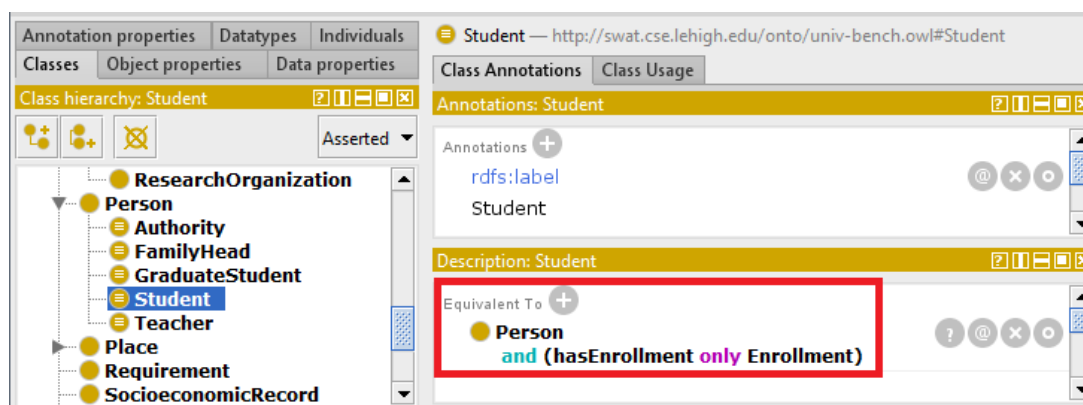


Figura 4.26: Condición de equivalencia definida en la clase *Student*

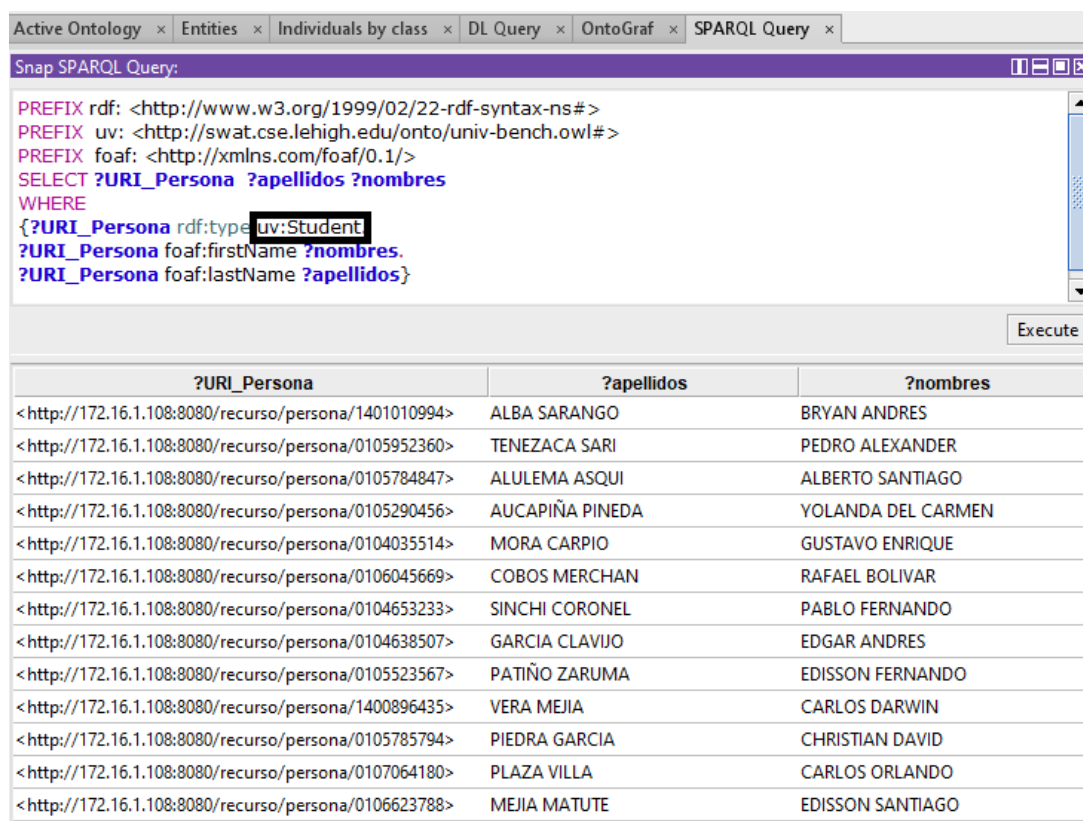
Antes de iniciar la clasificación de instancias a través de condiciones de equivalencia, en la figura 4.27 se presenta las instancias del tipo *Person* existentes en el archivo RDF.

Active Ontology x Entities x Individuals by class x DL Query x OntoGraf x SPARQL Query x		
Snap SPARQL Query:		
<pre> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> SELECT ?URI_Persona ?apellidos ?nombres WHERE { ?URI_Persona rdf:type uv:Person ?URI_Persona foaf:firstName ?nombres. ?URI_Persona foaf:lastName ?apellidos} </pre>		
Execute		
?URI_Persona	?apellidos	?nombres
<http://172.16.1.108:8080/recursos/persona/...	PATÍÑO UYAGUARI	NESTOR ARNOLDO
<http://172.16.1.108:8080/recursos/persona/...	CARPIO AGUDO	ESTHER ELENA
<http://172.16.1.108:8080/recursos/persona/...	GARCIA SEMPETEGUI	EDGAR GILBERTO
<http://172.16.1.108:8080/recursos/persona/...	ALBA SARANGO	BRYAN ANDRES
<http://172.16.1.108:8080/recursos/persona/...	GUEVARA TOLEDO	ANDREA PATRICIA
<http://172.16.1.108:8080/recursos/persona/...	JARAMILLO PALACIOS	FABIAN LEONARDO
<http://172.16.1.108:8080/recursos/persona/...	TENEZACA SARI	PEDRO ALEXANDER
<http://172.16.1.108:8080/recursos/persona/...	CORONEL SANTACRUZ	LUIS MODESTO
<http://172.16.1.108:8080/recursos/persona/...	VERA MOSQUERA	CARLOS ARIOSTO
<http://172.16.1.108:8080/recursos/persona/...	SARANGO MASACHE	BETTY JUDIT
<http://172.16.1.108:8080/recursos/persona/...	ALULEMA ASQUI	ALBERTO SANTIAGO
<http://172.16.1.108:8080/recursos/persona/...	MEJÍA HERRERA	ANGEL LEONARDO
<http://172.16.1.108:8080/recursos/persona/...	MORA CARPIO	GUSTAVO ENRIQUE
<http://172.16.1.108:8080/recursos/persona/...	COBOS MERCHAN	RAFAEL BOLIVAR
<http://172.16.1.108:8080/recursos/persona/...	ALULEMA ALVARO	RAUL GILBERTO
<http://172.16.1.108:8080/recursos/persona/...	GARCIA CLAVIJO	EDGAR ANDRES
<http://172.16.1.108:8080/recursos/persona/...	VERA MEJIA	CARLOS DARWIN
<http://172.16.1.108:8080/recursos/persona/...	CAMPOS ARGUDO	KARINA PAMELA
<http://172.16.1.108:8080/recursos/persona/...	TENEZACA TACURI	ANGEL HIPÓLITO
<http://172.16.1.108:8080/recursos/persona/...	PIEDRA GARCIA	CHRISTIAN DAVID
<http://172.16.1.108:8080/recursos/persona/...	RODRÍGUEZ VÉLEZ	MARÍA CECILIA
<http://172.16.1.108:8080/recursos/persona/...	VAZQUEZ PATÍÑO	ANGEL OSWALDO
<http://172.16.1.108:8080/recursos/persona/...	COBOS ROMERO	BOLIVAR RAFAEL
<http://172.16.1.108:8080/recursos/persona/...	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN
<http://172.16.1.108:8080/recursos/persona/...	PLAZA MELGAR	JOSÉ SAMUEL
<http://172.16.1.108:8080/recursos/persona/...	SINCHI CORONEL	PABLO FERNANDO
<http://172.16.1.108:8080/recursos/persona/...	PATÍÑO ZARUMA	EDISSON FERNANDO
<http://172.16.1.108:8080/recursos/persona/...	PLAZA VILLA	CARLOS ORLANDO
<http://172.16.1.108:8080/recursos/persona/...	MEJIA MATUTE	EDISSON SANTIAGO

Figura 4.27: SPARQL - Personas

Para obtener las instancias que son del tipo *Student*, se procede a ejecutar la consulta definida en la figura 4.28, el cual mediante un razonador aplica las reglas definidas en el modelo ontológico y retorna las instancias que cumplen la condición

de equivalencia y por lo tanto son considerados como estudiantes.



The screenshot shows a web-based SPARQL query interface. At the top, there are tabs for 'Active Ontology', 'Entities', 'Individuals by class', 'DL Query', 'OntoGraf', and 'SPARQL Query'. The 'SPARQL Query' tab is active. Below the tabs, there is a text area for the query. The query is as follows:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?URI_Persona ?apellidos ?nombres
WHERE
{
  ?URI_Persona rdf:type uv:Student
  ?URI_Persona foaf:firstName ?nombres.
  ?URI_Persona foaf:lastName ?apellidos
}
```

Below the query text area, there is an 'Execute' button. The results of the query are displayed in a table with three columns: '?URI_Persona', '?apellidos', and '?nombres'. The table contains 15 rows of data, each representing a student.

?URI_Persona	?apellidos	?nombres
<http://172.16.1.108:8080/recurso/persona/1401010994>	ALBA SARANGO	BRYAN ANDRES
<http://172.16.1.108:8080/recurso/persona/0105952360>	TENEZACA SARI	PEDRO ALEXANDER
<http://172.16.1.108:8080/recurso/persona/0105784847>	ALULEMA ASQUI	ALBERTO SANTIAGO
<http://172.16.1.108:8080/recurso/persona/0105290456>	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN
<http://172.16.1.108:8080/recurso/persona/0104035514>	MORA CARPIO	GUSTAVO ENRIQUE
<http://172.16.1.108:8080/recurso/persona/0106045669>	COBOS MERCHANT	RAFAEL BOLIVAR
<http://172.16.1.108:8080/recurso/persona/0104653233>	SINCHI CORONEL	PABLO FERNANDO
<http://172.16.1.108:8080/recurso/persona/0104638507>	GARCIA CLAVIJO	EDGAR ANDRES
<http://172.16.1.108:8080/recurso/persona/0105523567>	PATÍÑO ZARUMA	EDISSON FERNANDO
<http://172.16.1.108:8080/recurso/persona/1400896435>	VERA MEJIA	CARLOS DARWIN
<http://172.16.1.108:8080/recurso/persona/0105785794>	PIEDRA GARCIA	CHRISTIAN DAVID
<http://172.16.1.108:8080/recurso/persona/0107064180>	PLAZA VILLA	CARLOS ORLANDO
<http://172.16.1.108:8080/recurso/persona/0106623788>	MEJIA MATUTE	EDISSON SANTIAGO

Figura 4.28: SPARQL - Estudiantes

Para los casos de docente (ver figura 4.29) y jefes de familia (ver figura 4.30) se aplica el mismo criterio. Es decir, existen condiciones de equivalencia que permiten inferir a que tipo corresponde determinada instancia.

Snap SPARQL Query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ls: <http://linkedscience.org/teach/ns#>
SELECT ?URI_Persona ?apellidos ?nombres
WHERE
{
  ?URI_Persona rdf:type ls:Teacher
  ?URI_Persona foaf:firstName ?nombres.
  ?URI_Persona foaf:lastName ?apellidos}

```

Execute

?URI_Persona	?apellidos	?nombres
<http://172.16.1.108:8080/recursos/persona/...>	GUEVARA TOLEDO	ANDREA PATRICIA
<http://172.16.1.108:8080/recursos/persona/...>	JARAMILLO PALACIOS	FABIAN LEONARDO
<http://172.16.1.108:8080/recursos/persona/...>	VAZQUEZ PATIÑO	ANGEL OSWALDO
<http://172.16.1.108:8080/recursos/persona/...>	CAMPOS ARGUDO	KARINA PAMELA

Figura 4.29: SPARQL - Profesores

Active Ontology x Entities x Individuals by class x DL Query x OntoGraf x SPARQL Query x

Snap SPARQL Query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ls: <http://linkedscience.org/teach/ns#>
PREFIX uc: <http://ucuenca.edu.ec/ontology/>
SELECT ?URI_Persona ?apellidos ?nombres
WHERE
{
  ?URI_Persona rdf:type uc:FamilyHead
  ?URI_Persona foaf:firstName ?nombres.
  ?URI_Persona foaf:lastName ?apellidos}

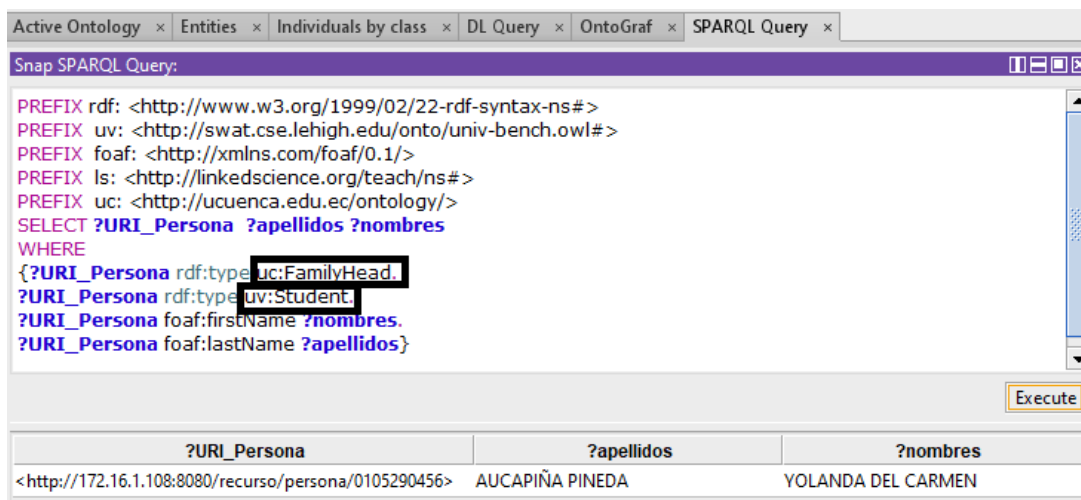
```

Execute

?URI_Persona	?apellidos	?nombres
<http://172.16.1.108:8080/recursos/persona/0102781978>	PATIÑO UYAGUARI	NESTOR ARNOLDO
<http://172.16.1.108:8080/recursos/persona/0102396421>	CARPIO AGUDO	ESTHER ELENA
<http://172.16.1.108:8080/recursos/persona/0102437738>	GARCIA SEMPERTEGUI	EDGAR GILBERTO
<http://172.16.1.108:8080/recursos/persona/0100016229>	CORONEL SANTACRUZ	LUIS MODESTO
<http://172.16.1.108:8080/recursos/persona/1400290647>	VERA MOSQUERA	CARLOS ARIOSTO
<http://172.16.1.108:8080/recursos/persona/1900318054>	SARANGO MASACHE	BETTY JUDIT
<http://172.16.1.108:8080/recursos/persona/0100050954>	COBOS ROMERO	BOLIVAR RAFAEL
<http://172.16.1.108:8080/recursos/persona/0105290456>	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN
<http://172.16.1.108:8080/recursos/persona/1102610829>	MEJÍA HERRERA	ANGEL LEONARDO
<http://172.16.1.108:8080/recursos/persona/0700817364>	PLAZA MELGAR	JOSÉ SAMUEL
<http://172.16.1.108:8080/recursos/persona/0601450596>	ALULEMA ALVARO	RAUL GILBERTO
<http://172.16.1.108:8080/recursos/persona/0101903508>	TENEZACA TACURI	ANGEL HIPÓLITO
<http://172.16.1.108:8080/recursos/persona/0100910421>	RODRÍGUEZ VÉLEZ	MARÍA CECILIA

Figura 4.30: SPARQL - Jefes Familia

Adicionalmente en la figura 4.31 se presenta una persona que es estudiante y jefe de familia a la vez.



Active Ontology x Entities x Individuals by class x DL Query x OntoGraf x SPARQL Query x

Snap SPARQL Query:

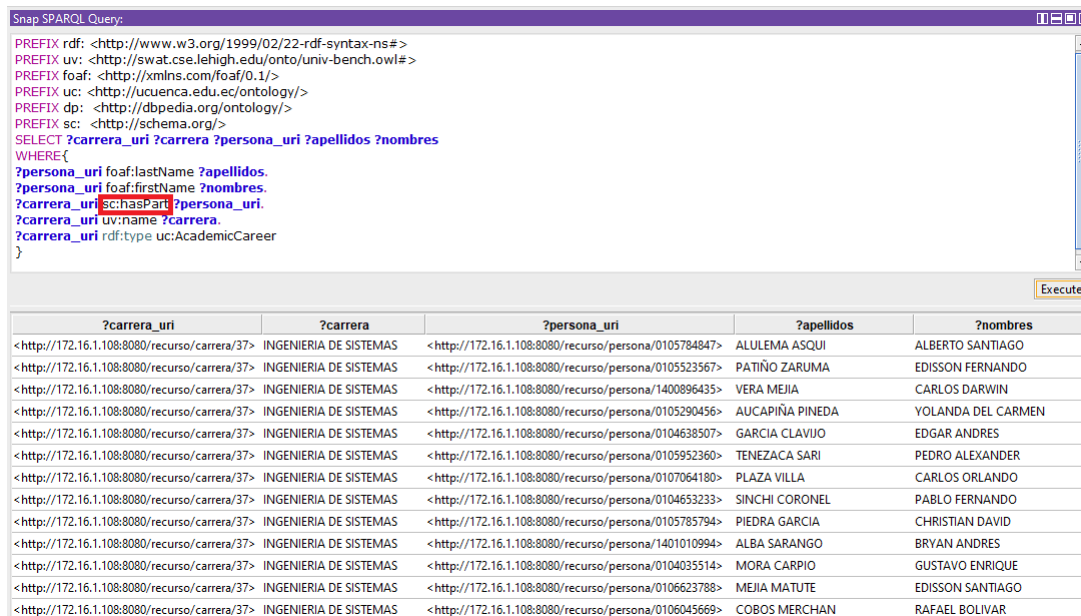
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ls: <http://linkedscience.org/teach/ns#>
PREFIX uc: <http://ucuenca.edu.ec/ontology/>
SELECT ?URI_Persona ?apellidos ?nombres
WHERE
{
  ?URI_Persona rdf:type uc:FamilyHead.
  ?URI_Persona rdf:type uv:Student.
  ?URI_Persona foaf:firstName ?nombres.
  ?URI_Persona foaf:lastName ?apellidos
}
```

Execute

?URI_Persona	?apellidos	?nombres
<http://172.16.1.108:8080/recurso/persona/0105290456>	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN

Figura 4.31: SPARQL - Estudiante y Jefe Familia

Además, la definición de relaciones inversas en el modelo ontológico permite la navegación desde ambos lados de la tripleta. Por ejemplo, en la consulta SPARQL de la figura 4.32 se extrae los estudiantes que forman parte de una carrera a través de la propiedad de objeto *hasPart*.



Snap SPARQL Query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX uc: <http://ucuenca.edu.ec/ontology/>
PREFIX dp: <http://dbpedia.org/ontology/>
PREFIX sc: <http://schema.org/>
SELECT ?carrera_uri ?carrera ?persona_uri ?apellidos ?nombres
WHERE{
  ?persona_uri foaf:lastName ?apellidos.
  ?persona_uri foaf:firstName ?nombres.
  ?carrera_uri sc:hasPart ?persona_uri.
  ?carrera_uri uv:name ?carrera.
  ?carrera_uri rdf:type uc:AcademicCareer
}
```

Execute

?carrera_uri	?carrera	?persona_uri	?apellidos	?nombres
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0105784847>	ALULEMA ASQUI	ALBERTO SANTIAGO
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0105523567>	PATIÑO ZARUMA	EDISSON FERNANDO
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/1400896435>	VERA MEJIA	CARLOS DARWIN
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0105290456>	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0104638507>	GARCIA CLAVIJO	EDGAR ANDRES
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0105952360>	TENEZACA SARI	PEDRO ALEXANDER
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0107064180>	PLAZA VILLA	CARLOS ORLANDO
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0104653233>	SINCHI CORONEL	PABLO FERNANDO
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0105785794>	PIEDRA GARCIA	CHRISTIAN DAVID
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0107064180>	ALBA SARANGO	BRYAN ANDRES
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0104035514>	MORA CARPIO	GUSTAVO ENRIQUE
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0106623788>	MEJIA MATUTE	EDISSON SANTIAGO
<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS	<http://172.16.1.108:8080/recurso/persona/0106045669>	COBOS MERCHANT	RAFAEL BOLIVAR

Figura 4.32: SPARQL - Estudiantes de carrera X

Con el objetivo de presentar el funcionamiento de propiedades inversas se obtendrá una salida similar al de la figura 4.32 utilizando su inversa (*isPartOf*). La consulta y el resultado se presenta en la figura 4.33.

Snap SPARQL Query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX uv: <http://swat.cse.lehigh.edu/onto/univ-bench.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX uc: <http://ucuenca.edu.ec/ontology/>
PREFIX dp: <http://dbpedia.org/ontology/>
SELECT ?persona_uri ?apellidos ?nombres ?carrera_uri ?carrera
WHERE {
  ?persona_uri foaf:lastName ?apellidos.
  ?persona_uri foaf:firstName ?nombres.
  ?persona_uri dp:isPartOf ?carrera_uri.
  ?carrera_uri uv:name ?carrera.
  ?carrera_uri rdf:type uc:AcademicCareer
}

```

?persona_uri	?apellidos	?nombres	?carrera_uri	?carrera
<http://172.16.1.108:8080/recurso/persona/ ...>	ALULEMA ASQUI	ALBERTO SANTIAGO	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/...>	PATINO ZARUMA	EDISSON FERNANDO	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	VERA MEJIA	CARLOS DARWIN	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	AUCAPIÑA PINEDA	YOLANDA DEL CARMEN	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	GARCIA CLAVIJO	EDGAR ANDRES	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	TENEZACA SARI	PEDRO ALEXANDER	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	PLAZA VILLA	CARLOS ORLANDO	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	SINCHI CORONEL	PABLO FERNANDO	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	PIEDRA GARCIA	CHRISTIAN DAVID	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	ALBA SARANGO	BRYAN ANDRES	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	MORA CARPIO	GUSTAVO ENRIQUE	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	MEJIA MATUTE	EDISSON SANTIAGO	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS
<http://172.16.1.108:8080/recurso/persona/>	COBOS MERCHANT	RAFAEL BOLIVAR	<http://172.16.1.108:8080/recurso/carrera/37>	INGENIERIA DE SISTEMAS

Figura 4.33: SPARQL - Carrera X de estudiantes

Capítulo 5

Prototipo del buscador

Con la finalidad de proveer al usuario una interfaz gráfica que permita realizar búsquedas sobre el modelo de integración de datos RDF, se presenta un prototipo de buscador semántico, que actúa sobre los servicios de Fuseki y ELDA levantados durante las etapas de publicación y explotación.

5.1. Descripción general

Una vez integrado los datos provenientes de las fuentes heterogéneas es necesario proveer al usuario de una interfaz que permita obtener información de forma sencilla y amigable, para esto se ha ajustado un prototipo de buscador semántico que toma como base el buscador semántico del proyecto FedQuest¹. A continuación, se describe los ajustes necesarios para su despliegue y configuración.

Proceso de despliegue del buscador semántico:

1. Clonar el proyecto FedQuest de GitHub².
2. Instalar el framework Meteor³ para desplegar el buscador.
3. Situar en el proyecto Fedquest, tal como se presenta en la figura 5.1.

¹FedQuest: Aplicación web para generar consultas federadas SPARQL utilizando una interfaz visual. Disponible en: <https://github.com/joe-programador/sparql-fedquest>

²GitHub: Plataforma de desarrollo colaborativo de software que permite el alojamiento de proyectos.

³Meteor: Plataforma para crear aplicaciones web en tiempo real. Permite sincronizar la base de datos de la aplicación y la interfaz de usuario.

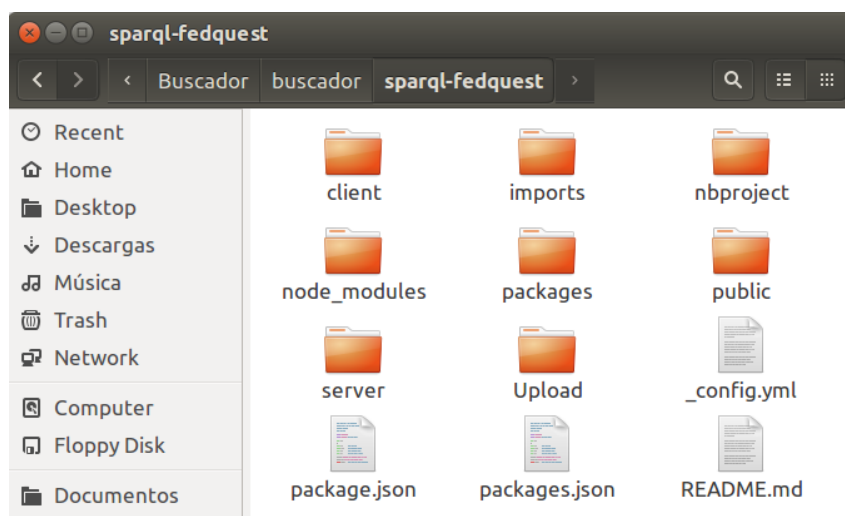


Figura 5.1: Proyecto - buscador Fedquest

4. Debido a que la solicitud de recursos (imágenes, etc) son redireccionados a una URL específica, es necesario declarar la ruta del servidor donde se desplegará el buscador semántico, esto se realiza por medio del comando *export ROOT_URL=http://172.16.1.108:3000*.
5. El despliegue del buscador semántico finalmente se realiza a través del comando *"meteor"*

Proceso de configuración del buscador semántico:

1. Registrar el endpoint desde «Constructor de consultas/Nuevo SPARQL Endpoint». Los parámetros de configuración a ingresar son la URL tanto del endpoint como del esquema. En la figura 5.2 se presenta la definición de tales parámetros.

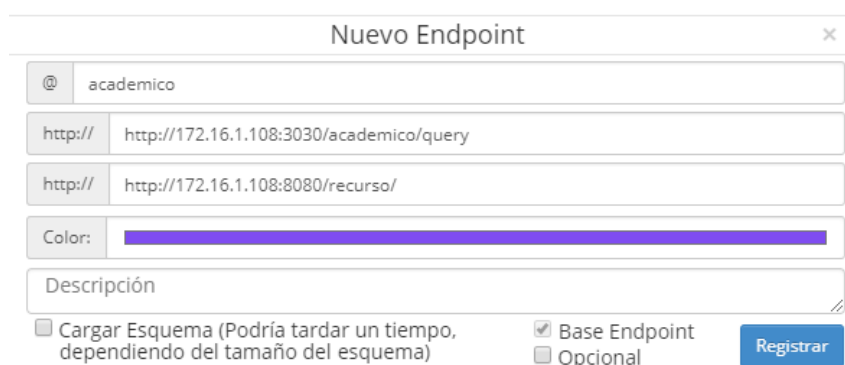


Figura 5.2: Registro Endpoint

2. Definir y configurar las entidades que serán presentadas por el buscador, esto se realiza desde la opción «Administrar/Configuración». En la figura 5.3 se presenta la configuración de la entidad Persona, en el cual se especifica los siguiente parámetros:

- Nombre: Nombre de la entidad. Se recomienda asignar un nombre descriptivo.
- URI: Corresponde a la URI de la entidad. Esta URI será la que se definió en el modelo ontológico.
- Ícono: Ícono que identificará a la entidad.
- Atributo identificativo: Propiedad de dato que resaltará sobre el resto de propiedades ante un resultado de búsqueda.
- Autocompletado: Propiedad de dato que se autocompletará cuando el usuario realice una búsqueda por este tipo de entidad.
- Atributo indexado: Propiedad de dato que se indexó en la etapa de Publicación (ver figura 4.21).



Figura 5.3: Definición y configuración de entidades

3. Establecer entidades que para la visualización gráfica. En la figura figura 5.4 se presenta la selección de entidades que se presentarán en la vista del tipo gráfico.

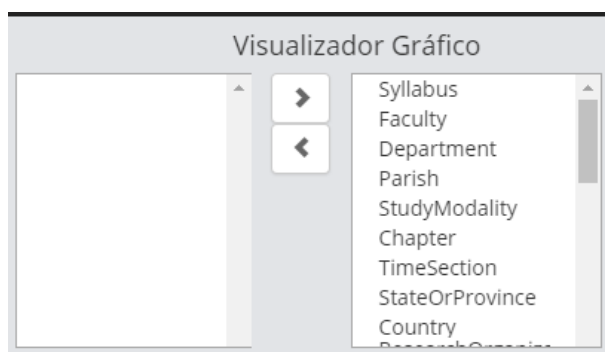


Figura 5.4: Selección de entidades - vista en modo grafo

4. Establecer entidades para filtrado. En esta etapa se definirá las entidades por las que el usuario filtrará la información (Ver figura 5.5). Estos filtros se incluirán automáticamente en la parte inferior de la barra de búsqueda una vez se seleccionen las entidades a filtrar.

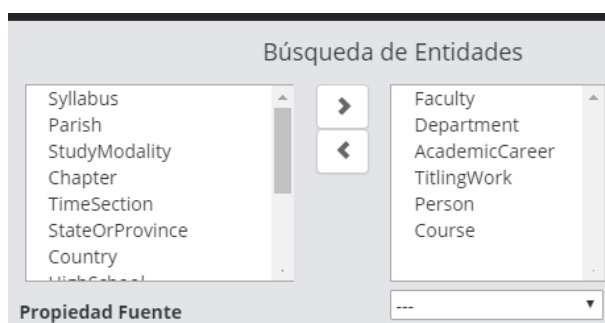


Figura 5.5: Selección entidades - filtros

5.2. Funcionamiento

El prototipo de buscador semántico despliega una interfaz intuitiva para el usuario, en la figura 5.6 se presenta la vista inicial del buscador. Para realizar una búsqueda el usuario deberá ingresar el texto que desea buscar en el campo "Entrada de texto". Adicionalmente se podrá aplicar determinados criterios de filtrado según como se definió en el paso 9. Los resultados que coincidan con el patrón de búsqueda serán listados en una nueva ventana, donde el usuario podrá visualizar esta información en modo de grafo o textual. Además, es posible obtener nueva información debido a que esta se encuentra enlazada.



Figura 5.6: Buscador semántico: Ventana Inicial

5.3. Presentación de resultados

Para la presentación de resultados, tal como se muestra en la figura 5.7, se trabajará con la Facultad de Ingeniería, por tanto se ingresará "*Facultad de Ingeniería*" en la entrada de texto y se filtrará por facultad para una búsqueda más precisa y limitar la cantidad de resultados al mismo tiempo.

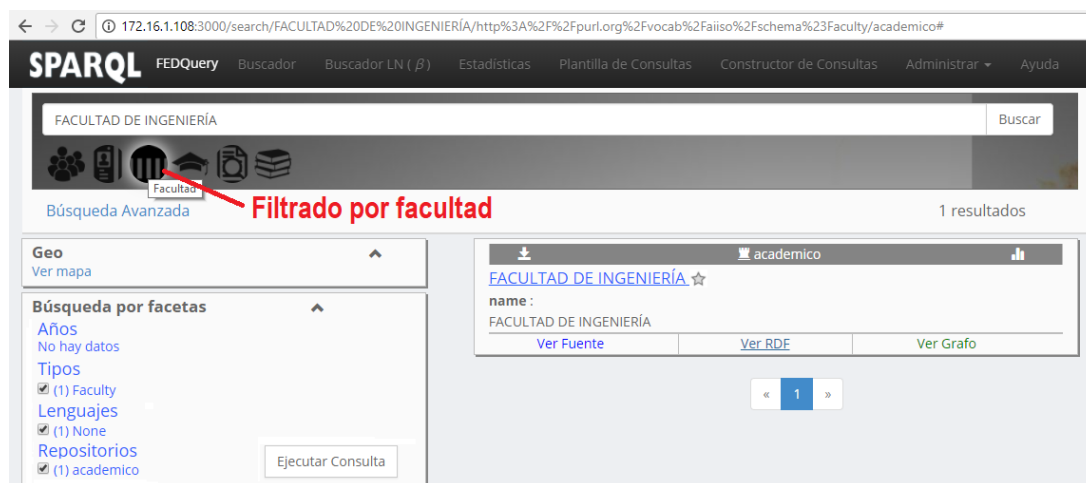
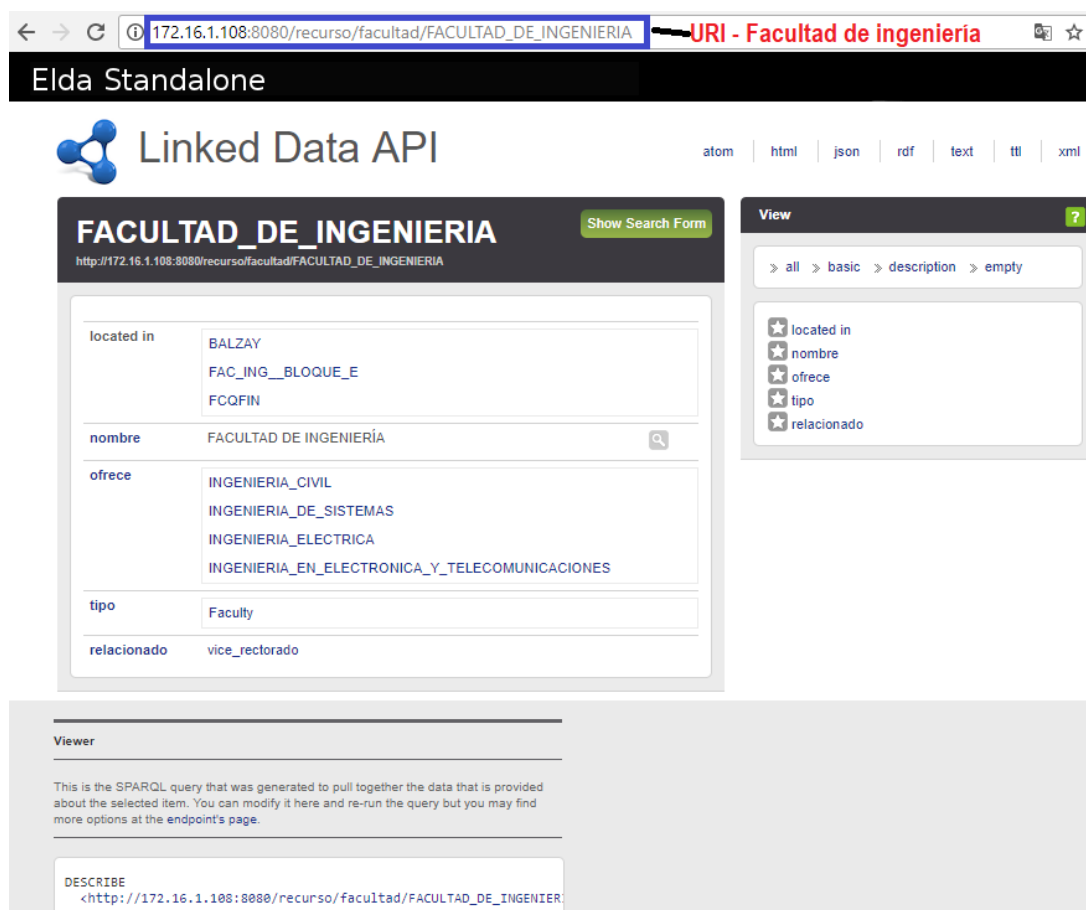


Figura 5.7: Resultados tras buscar "Facultad de Ingeniería"

Para la visualización de información detallada respecto a una determinada entidad, se

selecciona la opción "*Ver RDF*" (figura 5.7), esto genera una solicitud HTTP al servicio Elda mediante la URI del recurso generado para la Facultad de Ingeniería, el resultado se presenta en la figura 5.8. Adicionalmente, es posible visualizar información de los recursos que se encuentran enlazados mediante enlaces similares a los utilizados en una página HTML.



The screenshot displays a web browser window with the address bar showing the URI `172.16.1.108:8080/recurso/facultad/FACULTAD_DE_INGENIERIA`. The page title is "URI - Facultad de ingeniería". The main content area is titled "Linked Data API" and shows a "Show Search Form" button. Below this, the "FACULTAD_DE_INGENIERIA" resource is displayed with the following properties:

- located in:** BALZAY, FAC_ING_BLOQUE_E, FCQFIN
- nombre:** FACULTAD DE INGENIERÍA
- ofrece:** INGENIERIA_CIVIL, INGENIERIA_DE_SISTEMAS, INGENIERIA_ELECTRICA, INGENIERIA_EN_ELECTRONICA_Y_TELECOMUNICACIONES
- tipo:** Faculty
- relacionado:** vice_rectorado

On the right side, a "View" panel shows a list of properties: located in, nombre, ofrece, tipo, and relacionado. At the bottom, a "Viewer" section displays a SPARQL query that was generated to pull together the data for the selected item:

```
DESCRIBE
<http://172.16.1.108:8080/recurso/facultad/FACULTAD_DE_INGENIERIA>
```

Figura 5.8: Visualización de información relacionada a "Facultad de Ingeniería"

Para visualizar la visualización gráfica de la información respecto a una determinada entidad, se selecciona la opción "*Ver Grafo*" (figura 5.7). En la figura 5.9 se presenta el nodo generado para representar a la Facultad de Ingeniería.



Figura 5.9: Resultados desplegados en modo grafo - Facultad de Ingeniería

A diferencia del método anterior, la vista gráfica permite expandir el nodo inicial y presentar más información de la que inicialmente se proporciona. En la figura 5.10 se presenta los recursos enlazados a la facultad de ingeniería tras expandir el nodo de la figura 5.9. En este tipo de vista es posible distinguir cada una de las entidades que se encuentran enlazadas al recurso solicitado de acuerdo al ícono definido en la configuración de entidades.

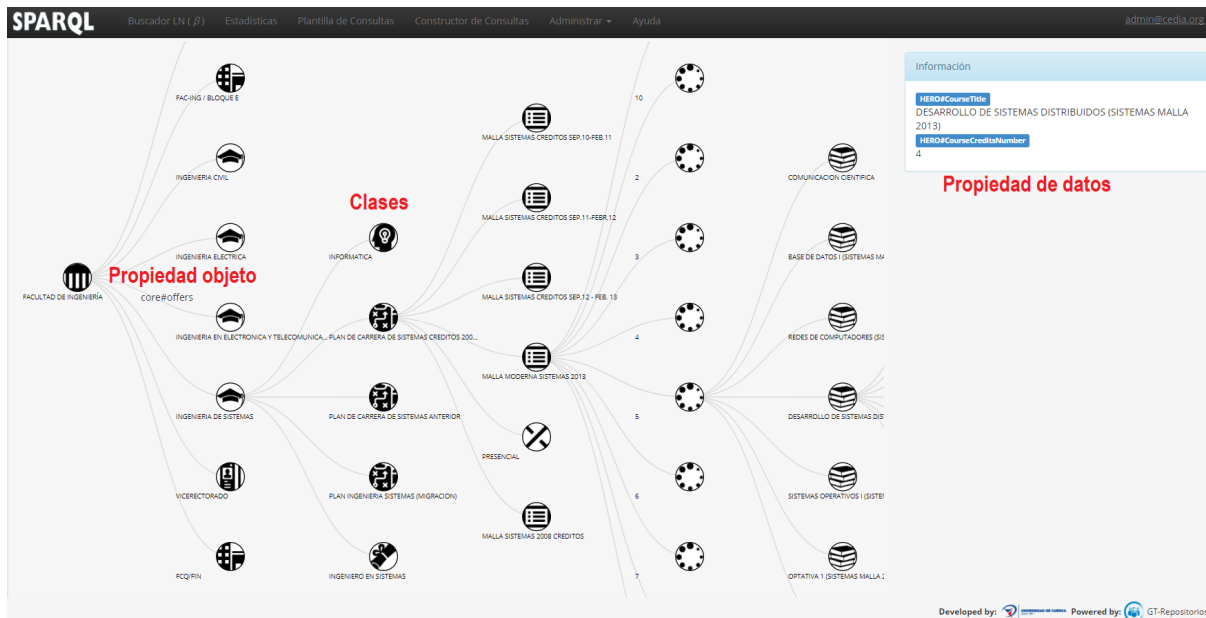


Figura 5.10: Expansión del grafo - Facultad de Ingeniería

Capítulo 6

Conclusiones y trabajos futuros

6.1. Conclusiones finales

Como se puede observar la ejecución en cascada de las diferentes actividades, establecidas en el ciclo de vida para datos enlazados propuesto por Villazón - Terrazas, ha permitido alcanzar el objetivo primordial de este proyecto de investigación de integrar los datos existentes en las diversas fuentes de datos heterogéneas en un único modelo de datos, para posteriormente permitir al usuario realizar búsquedas específicas mediante un buscador semántico.

El desarrollo de la red ontológica UCuenca-Ontology que describe la estructura organizacional y académica de la Universidad de Cuenca, además de hacer posible la integración de datos, permite reducir la ambigüedad de la información debido a que esta se encuentra anotada haciendo uso de conceptos bien definidos. Es importante destacar también que el disponer de un vocabulario desarrollado específicamente para el dominio de una universidad ecuatoriana, crea la posibilidad de que dicho vocabulario sea reutilizado por parte de otras universidades, facilitando de esta manera la comunicación e intercambio de información entre ellas.

Si bien la red ontológica obtenida como resultado del proceso de modelado en base a la metodología NeOn, no contempla todas las áreas existentes en el dominio universitario, define una ontología básica que puede ser extendida en proyectos posteriores, mediante la integración de ontologías desarrolladas para modelar áreas específicas de la universidad. La ampliación de la red ontológica implica necesariamente realizar actividades de generación y publicación con la finalidad de que los datos referentes a los nuevos conceptos y propiedades incluidas estén disponibles para la ejecución de búsquedas específicas.

Se utilizó la metodología de Villazon-Terrazas desde la especificación hasta la explotación con herramientas que automatizan cada una de estas etapas, por tanto el tiempo y la intervención humana se reduce drásticamente.

La implementación del buscador semántico universitario permite buscar información sin la necesidad de conocer lenguajes de consulta estándar, ni escribir consultas complejas, por tanto esa actividad es transparente para el usuario. Además, al actuar sobre información enlazada e integrada en un único repositorio, permite descubrir enlaces entre datos que respecto a una base de datos relacional no se puede determinar a simple vista.

El buscador es un proyecto web que cuenta con una interfaz similar a los principales buscadores web, por tanto, realizar búsquedas de información será intuitivo y amigable para el usuario.

6.2. Líneas de trabajos futuros

Durante el desarrollo del presente proyecto de investigación se han planteado temas abiertos que pueden ser considerados para trabajos futuros.

- **Ampliación de la red ontológica:** Agregar conceptos que describan áreas que no han sido consideradas en este proyecto. Para esto es necesario realizar el proceso descrito por la metodología NeOn con la finalidad de desarrollar una ontología que modele el área universitaria considerada y finalmente integrarla a la red ontológica.
- **Localización ontológica:** Si bien en el presente proyecto se agregaron anotaciones en lenguaje español a conceptos considerados relevantes, lo ideal es disponer de un vocabulario completamente en español ya que este se orienta hacia universidades ecuatorianas donde predomina dicho lenguaje.
- **Seguridad en el buscador:** Actualmente el buscador no cuenta con control de acceso, por tanto la información puede ser accedida por cualquier usuario que posea la dirección del servidor. Como trabajo futuro se plantea, dotar al buscador con características de seguridad como la autenticación y autorización de la información al cual el usuario puede acceder.
- **Aplicación del proceso en otros dominios:** Este proyecto plantea una idea general: integrar información de fuentes heterogéneas en un modelo único. Actualmente, no son las universidades las únicas instituciones que se ven afectadas por la incapacidad de realizar consultas sobre la totalidad de los datos existentes de forma rápida y sencilla, es por esto que el proceso descrito puede ser replicado para empresas, instituciones públicas y privadas, etc.

Anexo A

Documento de especificación de requerimientos ontológicos

1	Propósito
Proveer un modelo genérico de conocimiento que permita describir y presentar conceptos y relaciones existentes en el dominio universitario. Dicho modelo debe ser capaz de responder a consultas referentes a la estructura organizacional y académica de la Universidad de Cuenca. Adicionalmente el modelo obtenido debe almacenar la información necesaria para ser utilizado posteriormente en un sistema de búsqueda semántica.	
2	Alcance
La ontología plantea modelar la estructura organizacional y académica de la Universidad de Cuenca, para lo cual es necesario el almacenamiento de información de diversa índole dentro del contexto universitario. Dicha información se plantea en base a: <ul style="list-style-type: none">▪ Estructura organizacional: Organigrama de la estructura orgánica funcional y Estatuto de la Universidad de Cuenca▪ Estructura académica: Organigrama planteado posterior al estudio de los sistemas y fuentes de datos disponibles en la Universidad de Cuenca (Figura 4.1).	
3	Lenguaje de implementación
El lenguaje a utilizar para la implementación de la ontología es OWL/XML	
4	Usuarios finales previstos



<ul style="list-style-type: none">■ Usuario 1: Autoridades■ Usuario 2: Personal docente y administrativo■ Usuario 2: Estudiantes		
5	Usos previstos	
<ul style="list-style-type: none">■ Uso 1: Consultar información relacionada a la estructura organizacional■ Uso 2: Consultar información relacionada a la estructura académica■ Uso 3: Permitir conexión con modelos universitarios heterogéneos de áreas complementarias.		
6	Requerimientos de la ontología	
a. Requerimientos no funcionales		
b. Requerimientos funcionales – Preguntas de Competencia		
ID.	Pregunta de competencia (CQ)	Respuesta
Grupo 1: Preguntas relacionadas con organización administrativa		
PC1	¿Qué organismo es la máxima autoridad de la Universidad de Cuenca?	Consejo universitario
PC2	¿Quiénes son las principales autoridades de la Universidad de Cuenca?	Rector y vicerrector
PC3	¿Qué direcciones /unidades /coordinaciones están directamente relacionadas con el organismo X?	Dirección / Unidad / Coordinación 1 Dirección / Unidad / Coordinación 2
PC4	¿Qué unidades /coordinaciones incluye el organismo X?	Unidad / Coordinación 1 Unidad / Coordinación 2



PC5	¿Qué unidades incluye el organismo X perteneciente al organismo Y?	Unidad 1, Unidad 2,...
PC6	¿Qué unidades /departamentos /grupos de investigación pertenecen a la dirección de investigación?	Unidad /Departamento /Grupo de investigación 1 Unidad /Departamento /Grupo de investigación 2
PC7	¿Cuántas facultades tiene la Universidad de Cuenca?	12
PC8	¿Cuáles son las facultades de la Universidad de Cuenca?	Facultad 1, Facultad 2, Facultad 3,...
PC9	¿Qué carreras oferta la facultad X?	Carrera 1, Carrera 2, Carrera 3,...
Grupo 2: Preguntas relacionadas con organización académica		
PC10	¿Cuál es el plan de carrera actual de la carrera X?	Plan de carrera 1
PC11	¿Qué título universitario otorga la carrera X?	Título 1
PC12	¿Cuál es el área de conocimiento de la carrera X?	Área de conocimiento 1
PC13	¿Cuál es la modalidad de estudios de la carrera X?	Presencial /semipresencial /Distancia
PC14	¿Cuál es el número total de ciclos que tiene la carrera X?	10
PC15	¿Cuál es el número total de créditos /horas de la carrera?	255
PC16	¿Cuál es la malla curricular de la carrera X?	Malla curricular 1
PC17	¿Cuáles son las asignaturas de la malla curricular X?	Asignatura 1, Asignatura 2,...
PC18	¿Cuántas asignaturas tiene la malla curricular X?	55
PC19	¿Cuál es el código de la asignatura X?	2262
PC20	¿Cuál es el número de créditos /horas de la asignatura X?	6
PC21	¿Cuál es el eje de formación de la asignatura X?	Eje de formación 1



PC22	¿En qué nivel se dicta la asignatura X?	1, 2, 3,... 10
PC23	¿Cuáles son los prerrequisitos de la asignatura X?	Asignatura 1, Asignatura 2,...
PC24	¿Cuáles son los correquisitos de la asignatura X?	Asignatura 1, Asignatura 2,...
PC25	¿A qué asignaturas es equivalente la asignatura X?	Asignatura 3 Asignatura 4 + Asignatura 5
PC26	¿En qué carreras se oferta la asignatura X?	Carrera 1, Carrera 2, Carrera 3,...
PC27	¿Cuál es el período académico vigente en la carrera X?	Mes año inicio – Mes año fin
PC28	¿Cuántas asignaturas dicta el docente?	1, 2, 3,...
PC29	¿Cuántas horas de clase dicta el docente?	10, 12,...
PC30	¿En qué carreras dicta clases el docente?	Carrera 1, carrera 2, carrera 3,...
PC31	¿Qué docente es responsable de asignatura X durante un período académico Y?	Docente 1, docente 2...
PC32	¿Cuál es el nombre/número del grupo ofertado para la asignatura X durante el período académico Y?	1, 2... A,B...
PC33	¿Qué contenido tiene la asignatura X durante el período académico Y?	Contenido 1, contenido2,...
PC34	¿Qué criterios de evaluación se aplicarán en la asignatura X durante el período académico Y?	Trabajos individuales / Pruebas / Exámenes
PC35	¿Qué valor tiene el criterio de evaluación X de la asignatura Y durante el período académico Z?	60 / 20 / 20
PC36	¿Qué asignaturas se ofertan durante el período académico X en la carrera Y?	Asignatura 1, Asignatura 2, Asignatura 3,...



PC37	¿Cuál es el cupo máximo del grupo X en la asignatura Y durante el período académico Z?	30
PC38	¿Cuál es el cupo mínimo del grupo X en la asignatura Y durante el período académico Z?	10
PC39	¿Cuál es la sección del grupo X en la asignatura Y durante el período académico Z?	Matutina /Vespertina /Nocturna
PC40	¿En qué horario se oferta la asignatura X durante el período académico Y?	Día – Hora inicio – Hora Fin
PC41	¿En qué aula se oferta la asignatura X durante el período académico Y?	Aula 101, Aula 303,...
PC42	¿En qué período académico registra matrícula el estudiante X?	Período académico 1, período académico 2,...
PC43	¿En qué carrera registra matrícula el estudiante X?	Carrera 1, Carrera 2,...
PC44	¿En qué asignaturas registra matrícula el estudiante X durante el período académico Y?	Asignatura 1, Asignatura 2,...
PC45	¿Cuál es el estado de matrícula del estudiante X en la asignatura Y durante el período académico Z?	Matriculado – Prematriculado - Anulado
PC46	¿Cuál es el número de veces que el estudiante X repite la asignatura Y?	1, 2, 3,...
PC47	¿Cuál es el horario del estudiante X durante el período académico Y?	Asignatura 1 – Horario 1 – Grupo 1 - Espacio físico 1 Asignatura 2 – Horario 2 - Grupo 1 – Espacio físico 2
PC48	¿Cuál es el estado de aprobación del estudiante X en la asignatura Y?	Aprobado / Reprobado / Suspenso
PC49	¿Cuál es la forma de aprobación del estudiante X en la asignatura Y?	Homologación / Escolaridad



PC50	¿Cuál es la nota final del estudiante X en la asignatura Y?	80, 60,...
Grupo 3: Preguntas relacionadas con espacios físicos		
PC51	¿Cuáles son las coordenadas geográficas (latitud, longitud) del edificio donde se localiza la facultad X?	(2.9000, 2.8890)
PC52	¿Cuántos pisos tiene el edificio donde se localiza la facultad X?	1, 2, 3...
PC53	¿Cuántas aulas de clases tiene el edificio donde se localiza la facultad X?	20, 30,...
PC54	¿Cuál es el número del piso del edificio X en el que se localiza el aula de clases Y?	1, 2, 3,...
PC55	¿De qué tipo es el aula de clases X?	Auditorio, Aula, Laboratorio, etc.
PC56	¿Cuál es la capacidad del aula de clases X?	70, 100,...
PC57	¿Cuál es el tamaño (área) del aula X?	Piso 1
Grupo 4: Preguntas relacionadas con trabajos de titulación		
PC58	¿Cuál es el título del trabajo de titulación del estudiante X?	Título 1
PC59	¿Qué estudiante o estudiantes son los autores del trabajo de titulación X?	Persona 1, Persona2,...
PC60	¿Qué docente es el director del trabajo de titulación X?	Persona 1
PC61	¿Cuál es la calificación del trabajo de titulación X?	90/100
PC62	¿Cuál es el url del trabajo de titulación X?	url1
Grupo 5: Preguntas relacionadas con datos personales		
PC63	¿En qué otra universidad estudió el estudiante X?	Universidad Y
PC64	¿El estudiante X ha cursado una carrera diferente?	Si / No



PC65	¿En qué colegio se graduó el estudiante X?	Colegio Y
PC66	¿Qué título de bachiller tiene el estudiante X?	Título de bachiller 1
PC67	¿Cuál es el año de graduación como bachiller del estudiante X?	2011
PC68	¿En qué períodos académicos registra ficha socioeconómica el estudiante X?	Periodo Académico 1, Periodo Académico 2,...
PC69	¿Cuál es la fecha de inicio de estudios del estudiante X?	05/09/2011
PC70	¿Cuál es la fecha de egreso del estudiante X?	20/02/2017
PC71	¿Cuál es la fecha de graduación del estudiante X?	22/09/2015
PC72	¿Cuál es el mecanismo de titulación del estudiante X?	Trabajo de titulación /Examen complejo /...
PC73	¿Cuál es la nota de graduación del estudiante X?	80/100
PC74	¿Cuál es el número de refrendación del acta de grado del estudiante X?	01078956
PC75	¿Cuál es el período académico durante el cual se graduó el estudiante X?	Periodo académico 1
PC76	¿Cuál es el título universitario obtenido por el estudiante X?	Título universitario 1
PC77	¿Cuál es la malla curricular de la que graduó el estudiante X?	Malla curricular 1
PC78	¿Cuál es el nombre del Docente / Estudiante / Jefe de familia X?	Persona 1 (Apellido, Nombre)
PC79	¿Cuál es el número de identificación del Docente / Estudiante X?	010856493
PC80	¿Cuál es el número telefónico convencional del estudiante / docente X?	072897735
PC81	¿Cuál es el número telefónico celular del Estudiante / Docente X?	0968884502



PC82	¿Cuál es la dirección de correo electrónico institucional del estudiante / docente X?	dirección de correo 1
PC83	¿Cuál es la dirección de correo electrónico personal del estudiante / docente X?	dirección de correo 2
PC84	¿Cuál es la nacionalidad del estudiante X?	Ecuador / Colombia / ...
PC85	¿Cuál es el lugar de nacimiento del estudiante X?	Ciudad – Cantón – Provincia – País
PC86	¿Cuál es la fecha de nacimiento del estudiante / jefe de familia X?	13/02/1984
PC87	¿Cuál es el estado civil del estudiante X?	Soltero / Casado / Viudo / Divorciado
PC88	¿Cuál es el género del estudiante X?	Masculino / Femenino
PC89	¿Cuál es el tipo de sangre del estudiante X?	ORH+ / BRH+ / ARH+ / ...
PC90	¿Qué lateralidad tiene el estudiante X?	Diestro / Zurdo
PC91	¿Cuál es la dirección del domicilio del estudiante X?	Lugar Residencia – Dirección – Número de casa
PC92	¿Cuál es la etnia con la que se identifica el estudiante X?	Mestizo / Blanco / ...
PC93	¿El estudiante tiene algún tipo de discapacidad X?	Si / No
PC94	¿Qué tipo de discapacidad tiene el estudiante X?	Visual / auditiva / lenguaje / física
PC95	¿El estudiante X tiene carné de discapacidad?	Si / No
PC96	¿Qué porcentaje de discapacidad tiene el estudiante X?	5, 10, 20...
PC97	¿Cuál es la ocupación del jefe de familia X?	Empleado / ...
Grupo 6: Preguntas relacionadas con datos socioeconómicos		
PC98	¿Qué tipo de tenencia tiene la vivienda familiar del estudiante X?	Arrendada / Cedida / ...



PC99	¿Qué material de construcción tiene la vivienda familiar del estudiante X?	Ladrillo o bloque /Adobe o similar /...
PC100	¿En qué zona de residencia se ubica la vivienda familiar del estudiante X?	Urbana / Rural
PC101	¿De qué tipo es la vivienda familiar del estudiante X?	Casa / Cuarto /...
PC102	¿Cuál es la ubicación de la vivienda familiar del estudiante X?	Cantón – Provincia - País
PC103	¿Cuál es la dirección de la vivienda familiar del estudiante X?	Dirección 1
PC104	¿Cuál es el pago mensual por arriendo o hipoteca de la vivienda familiar del estudiante X?	650,00
PC105	¿Cuál es el teléfono de la vivienda familiar del estudiante X?	2858469
PC106	¿Cuál es el nombre del jefe del grupo familiar del estudiante?	Persona 1 (Apellido, Nombre)
PC107	¿Cuál es el rol del jefe del grupo familiar del estudiante X?	Padre (madre) /...
PC108	¿Cuál es el número de miembros que tiene el grupo familiar del estudiante X?	2, 3, 4...
PC109	¿Cuántos miembros del grupo familiar del estudiante X estudian en niveles diferentes al superior?	0, 1, 2,...
PC110	¿Cuántos miembros del grupo familiar del estudiante X son hijos menores de 6 años?	0, 1, 2, 3,...
PC111	¿El estudiante X es carga familiar de un servidor de la universidad?	Si / No
PC112	¿Cuántas propiedades no renteras / renteras / vacacionales tiene la familia del estudiante X?	0, 1, 2,...
PC113	¿Cuántas líneas telefónicas tiene la familia del estudiante X?	0, 1, 2,...
PC114	¿Cuántos vehículos tiene la familia del estudiante X?	0, 1, 2,...



PC115	¿Cuál es el valor acumulado del avalúo de los vehículos que tiene la familia del estudiante X?	20 000,00
PC116	¿La familia del estudiante X dispone de servicio de TVCable o similares?	Si / No
PC117	¿Cuánto ingresa mensualmente al grupo familiar del estudiante X por sueldos / pensiones / rentas /...?	1200 / 0 / 0
PC118	¿Cuánto egresa mensualmente del grupo familiar del estudiante X por aporte al IEES / impuesto a la renta /...?	110 / 0 / 10

7. Pre-glosario de términos		
a. Términos de preguntas de competencia		
Nombres		
estudiante (67)	asignatura (29)	período académico (18)
carrera (14)	docente (11)	grupo familiar (11)
vivienda familiar (8)	aula de clases (7)	facultad (6)
organismo (5)	trabajo de titulación (5)	dirección /unidad /coordinación (4)
edificio (4)	grupo (4)	jefe de familia (4)
mallla curricular (4)	criterio de evaluación (3)	matrícula (3)
miembros que estudian en otros niveles (3)	autoridad (2)	dirección/departamento /grupo de investigación (2)
horario (2)	nivel (2)	piso (2)
teléfono (2)	tipo de discapacidad (2)	título universitario (2)
universidad (2)	acta de grado	año de graduación como bachiller
área de conocimiento	autor	avalúo de vehículos
calificación	carné de discapacidad	código
colegio	contenido	coordenadas geográficas
correquisito	cupo máximo	cupo mínimo
dirección	dirección de domicilio	director
egreso (gasto)	eje de formación	email institucional



email personal	es carga de un servidor universitario	estado civil
estado de aprobación	estado de matrícula	estudiantes de otros niveles
etnia	fecha de egreso	fecha de graduación
fecha de inicio de estudios	ficha socioeconómica	forma de aprobación
género	ingreso	lateralidad
latitud	longitud	lugar de nacimiento
material de construcción	mecanismo de titulación	menores de 6 años
modalidad de estudios	nacionalidad	nombre
nombre /número	nota de graduación	nota final
número de créditos /horas	número de horas de clase	número de identificación
número de líneas telefónicas	número de piso	número de propiedades renteras
número de propiedades no renteras	número de propiedades vacacionales	número de refrendación
número de veces que repite	número de vehículos	número total de ciclos
número total de créditos /horas	ocupación	pago mensual por arriendo o hipoteca
plan de carrera	porcentaje de discapacidad	prerrequisito
rol familiar	sección	tamaño (área)
teléfono celular	tipo de sangre	tipo de tenencia
tipo de vivienda	título (bibliográfico)	título de bachiller
TVCable o similar	url	valor
zona de residencia		
Verbos		
tener (8)	ofertar durante (7)	ofertar (6)
dictar (4)	localizar (4)	registrar (4)
graduar (2)	incluir (2)	pertenecer (2)
relacionar	es autoridad máxima	es autoridad principal
es equivalente a	es responsable de	está vigente
estudiar	obtener	otorgar
ubicar		
Objetos		



Universidad de Cuenca (4)	Sueldo /Pensión /Renta	Aporte al IEES /
b. Términos de respuestas		
Nombres		
apellido	cantón	ciudad
día de la semana	hora fin	hora inicio
nombre	país	provincia
Objetos		
consejo universitario	rector	vicerrector
presencial /semipresen- cial /distancia	auditorio /aula /labora- torio /	trabajos individuales /pruebas /examen
matutina /vespertina /nocturna	casa /cuarto	mestizo /blanco /mon- tubio
matriculado /pre- matriculado /anulado	homologación /escolari- dad	trabajo de titulación /examen complejo
soltero /casado /viudo	masculino /femenino	ladrillo o bloque /adobe o similar
arrendada /cedida	urbana /rural	visual /auditiva /len- guaje /física

Anexo B

Vocabularios reutilizados en la red ontológica

Prefijo	URI
URI base de UCuenca-Ontology	
uco	http://localhost:8080/ontology/
Vocabularios importados	
core	http://vivoweb.org/ontology/core#
hero	http://www.UniversityReferenceOntology.org/HERO#
univ-bench	http://swat.cse.lehigh.edu/onto/univ-bench.owl#
time	http://www.w3.org/2006/time#
Vocabularios reutilizados para incluir términos específicos	
aiiso	http://purl.org/vocab/aiiso/schema#
amigo	http://amigo.gforge.inria.fr/owl/ContextVocabulary.owl#
bibo	http://purl.org/ontology/bibo/
bio	http://purl.org/vocab/bio/0.1/
cwmo	http://purl.org/cwmo/#
dbpedia	http://dbpedia.org/ontology/
ebucore	http://www.ebu.ch/metadata/ontologies/ebucore/ebucore#
foaf	http://xmlns.com/foaf/0.1/
frapo	http://purl.org/cerif/frapo/
gci	http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl#
gem	http://purl.oclc.org/gem/gemtype/
geo	http://www.w3.org/2003/01/geo/wgs84_pos#
geopolitical	http://aims.fao.org/aos/geopolitical.owl#
gnd	http://d-nb.info/standards/elementset/gnd#



gr	http://purl.org/goodrelations/v1#
icontact	http://ontology.eil.utoronto.ca/icontact.owl#
km4city	http://www.disit.org/km4city/schema#
locah	http://data.archiveshub.ac.uk/def/
midas	http://www.lia.deis.unibo.it/research/MIDAS/Ontologies/MIDAS_Profile.owl#
obo	http://purl.obolibrary.org/obo/
ocw	http://simile.mit.edu/2006/01/ontologies/ocw#
person	https://www.w3.org/ns/person#
person-owl	http://ebiquity.umbc.edu/ontology/person.owl#
profile	http://daml.umbc.edu/ontologies/profile-ont#
proton	http://www.ontotext.com/proton/protontop#
rda	http://rdaregistry.info/Elements/u/
saref4ee	http://ontology.tno.nl/saref4ee/#
schema	http://schema.org/
seas	https://w3id.org/seas/
sem	http://semanticweb.cs.vu.nl/2009/11/sem/
swrc	http://swrc.ontoware.org/ontology
teach	http://linkedscience.org/teach/ns#
terms	http://purl.org/dc/terms/
trip-ont	http://www.daml.org/2001/06/expenses/trip-ont.daml#
vcard	http://www.w3.org/2006/vcard/ns#

Anexo C

Diagrama de la red ontológica

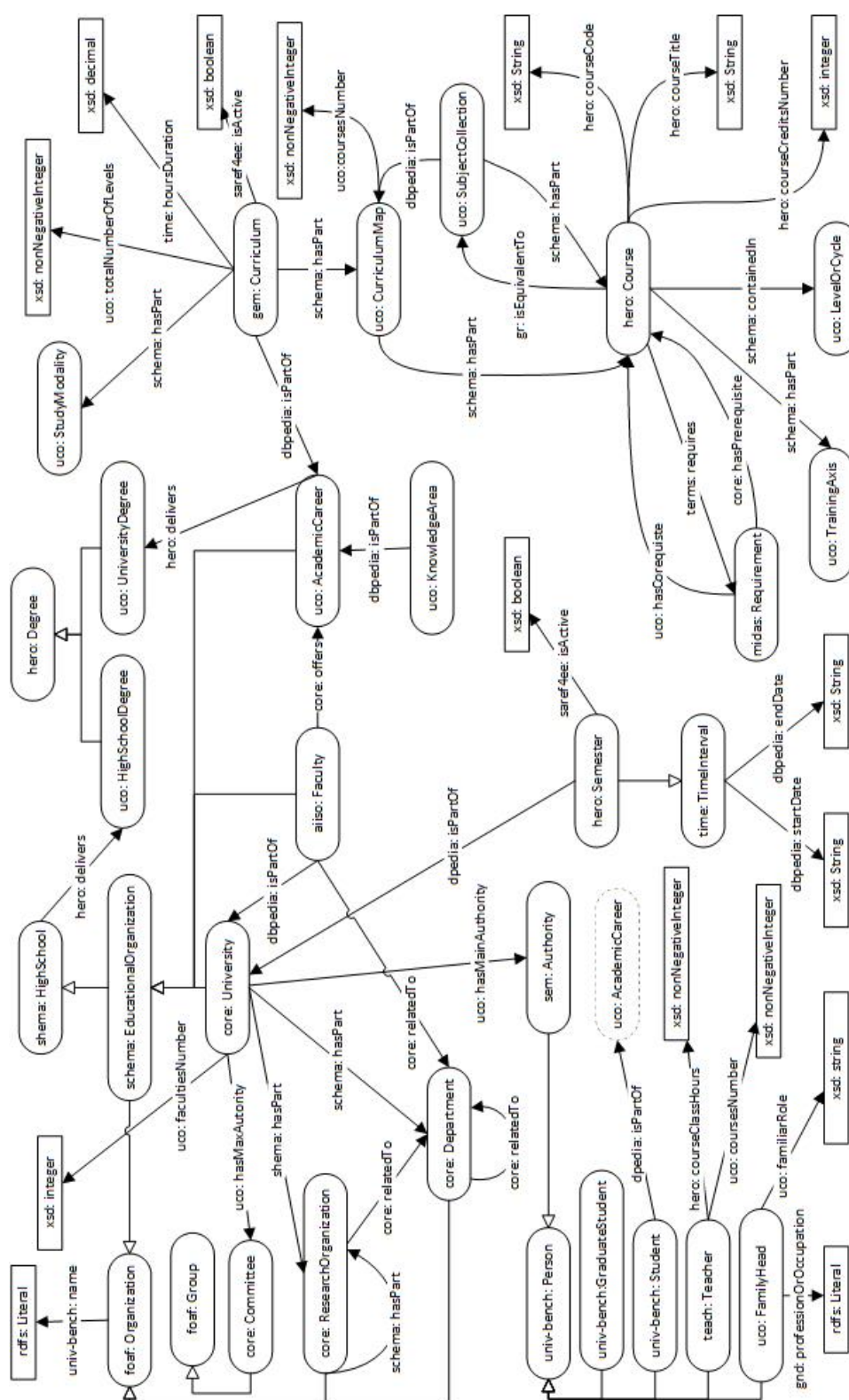


Figura C.1: Diagrama de la red ontológica: Parte 1

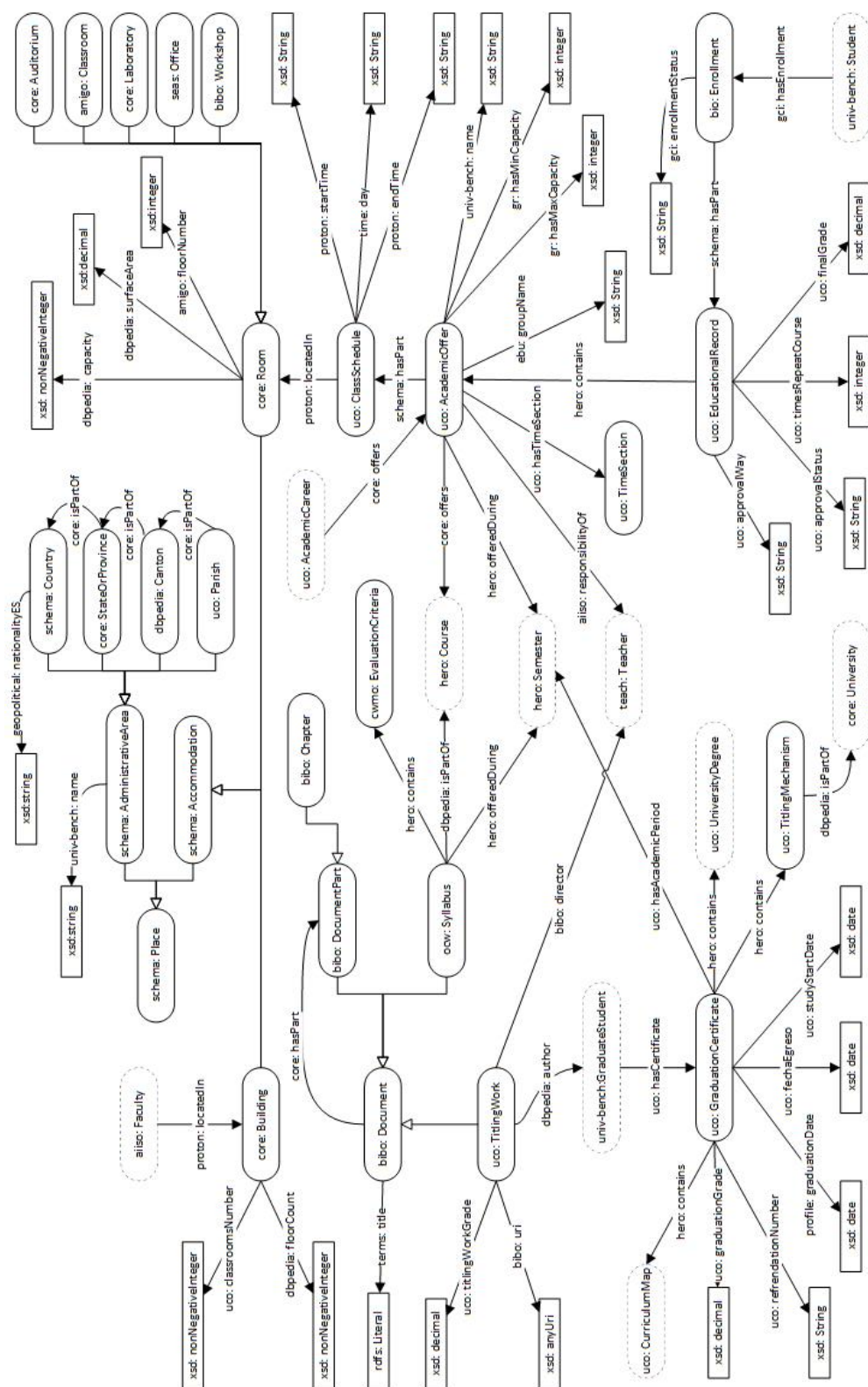


Figura C.2: Diagrama de la red ontológica: Parte 2

ANEXO C. DIAGRAMA DE LA RED ONTOLÓGICA

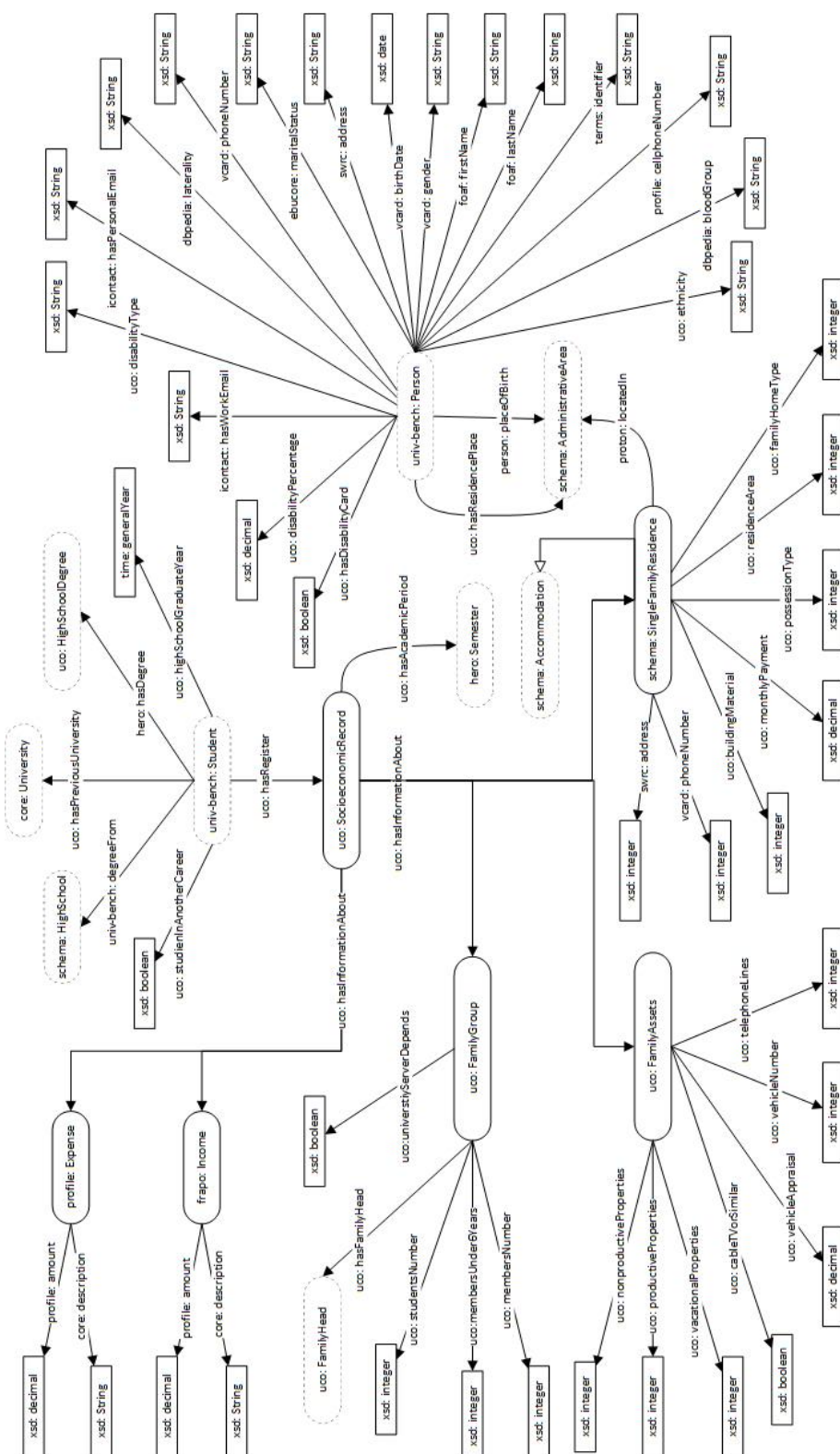


Figura C.3: Diagrama de la red ontológica: Parte 3

Anexo D

Clases definidas en el vocabulario UCuenca-Ontology

Prefijo	Clase	Término ORSD	Descripción
univ-bench	University	Universidad	Institución de educación superior.
aiiso	Faculty	Facultad	Sección de la universidad que comprende estudios en una rama específica.
uco	AcademicCareer	Carrera	Esquema de formación profesional para los estudiantes.
schema	HighSchool	Colegio	Permite representar el colegio en el cual un estudiante realizó sus estudios secundarios.
core	Committee	Consejo Universitario	Grupo de personas que conforman la máxima autoridad universitaria.
core	Department	Dirección, unidad y coordinación	Permite representar las diversas unidades organizacionales, sus relaciones y dependencias.
core	Research Organization	Dirección, unidad y grupo de investigación	Permite representar organismos destinados a labores de investigación, sus relaciones y dependencias.
sem	Authority	Autoridad	Personas elegidas como autoridades principales de la universidad.



Prefijo	Clase	Término ORSD	Descripción
univ-bench	Graduate Student	Estudiante graduado	Persona que ha concluido sus estudios, por lo tanto, registra un acta de grado y en algunos casos es autor de un trabajo de titulación.
uco	FamilyHead	Jefe de familia	Persona que trabaja y lleva su salario al hogar para mantener a su familia.
teach	Teacher	Docente	Permite identificar a docentes.
univ-bench	Student	Estudiante	Persona que registra matrícula en la universidad.
uco	HighSchoolDegree	Título de bachiller	Permite representar el título de bachiller otorgado por un colegio a un estudiante.
uco	UniversityDegree	Título universitario	Título otorgado a un estudiante al terminar sus estudios en una carrera universitaria.
ocw	Syllabus	Sílabo	Documento que contiene la planificación de una asignatura para un período específico.
uco	TitlingWork	Trabajo de titulación	Trabajo realizado por un estudiante para obtener un título universitario.
uco	AcademicOffer	Oferta académica	Permite representar qué asignaturas son ofertadas en un período determinado, su horario, docente responsable, nombre del grupo y número de cupos (mínimo y máximo).
uco	ClassSchedule	Horario de clases	Permite especificar el día, hora inicial, hora final y espacio físico de cada una de las ofertas académicas.
hero	Course	Asignatura	Permite representar una asignatura específica, sus requisitos, prerrequisitos y equivalencias.



Prefijo	Clase	Término ORSD	Descripción
gem	Curriculum	Plan de carrera	Permite representar la información general de la carrera e incluye la malla curricular vigente.
uco	CurriculumMap	MallaCurricular	Contiene todas las asignaturas que un estudiante debe cursar en una determinada carrera.
uco	Educational Record	Registro académico	Contiene información sobre las asignaturas en las cuales un estudiante se ha matriculado.
bio	Enrollment	Matrícula	Permite enlazar un estudiante con una oferta académica.
cwmo	EvaluationCriteria	Criterio de evaluación	Especifica los criterios que serán empleados para evaluar una asignatura determinada.
profile	Expense	Egreso	Permite representar el monto y descripción de los gastos registrados en una ficha socioeconómica.
uco	FamilyAssets	Bienes familiares	Contiene información sobre los bienes familiares registrados en una ficha socioeconómica.
uco	FamilyGroup	Grupo familiar	Contiene información sobre el grupo familiar registrada en una ficha socioeconómica.
uco	Graduation Certificate	Acta de grado	Permite representar el documento legal que avala que un estudiante ha culminado una carrera universitaria.
frapo	Income	Ingreso	Permite representar el monto y descripción de los ingresos registrados en una ficha socioeconómica.
uco	KnowledgeArea	Área de conocimiento	Ámbitos específicos de conocimiento, con su naturaleza particular y diferentes métodos para adquirir conocimientos.
uco	LevelOrCycle	Nivel	Permite representar los niveles en los cuales se organizan las asignaturas en una carrera.



Prefijo	Clase	Término ORSD	Descripción
uco	TitlingMechanism	Mecanismo de titulación	Modalidad definida que un estudiante egresado debe cumplir para obtener un título universitario.
midas	Requirement	Requisito	Permite representar prerrequisitos y correquisitos de una asignatura específica.
uco	Socioeconomic Record	Ficha socioeconómica	Contiene la información socioeconómica registrada por un estudiante previo a realizar un proceso de matrícula.
uco	StudyModality	Modalidad de estudios	Permite diferenciar la modalidad en que se imparten los estudios académicos en una carrera: presencial, semipresencial y distancia.
uco	SubjectCollection	Conjunto de asignaturas	Permite describir la equivalencia de una asignatura.
hero	Semester	Período académico	Representa un período académico que generalmente tiene una duración de seis meses.
uco	TimeSection	Sección	Permite diferenciar la sección en la cual se dicta una clase: matutina, vespertina o nocturna.
uco	TrainingAxis	Eje de formación	Permite diferenciar el tipo de asignatura: optativa, libre elección, etc.

Anexo E

Propiedades de objetos definidas en el vocabulario UCuenca-Ontology

Propiedad	Descripción	Inversa	Dominio	Rango
awardedBy	Relaciona un título con la organización educativa que lo otorga.	delivers	Degree	Educational Organization
contains	Relaciona un nivel con las asignaturas que contiene.	containedIn	Level	Course
graduateFrom	Relaciona un estudiante con el colegio del cual se graduó.	hasAlumnus	Student	HighSchool
hasAcademic Period	Permite establecer un período académico específico.	isAcademic PeriodOf	•Graduation Certificate •Socioeconomic Record	Semester
hasAuthor	Relaciona un trabajo de titulación con su autor.	isAuthorOf	TitlingWork	Graduate Student
hasCertificate	Relaciona un estudiante graduado con su acta de grado.	isCertificateOf	Graduate Student	Graduation Certificate
hasCorequisite	Permite establecer los correquisitos de una asignatura.	corequisiteFor	Requirement	Course



Propiedad	Descripción	Inversa	Dominio	Rango
hasDirector	Relaciona un trabajo de titulación con su director.	isDirectorOf	TitlingWork	Professor
hasEnrollment	Relaciona un estudiante con una matrícula.	isEnrollmentOf	Student	Enrollment
hasDegree	Relaciona un estudiante con su título de bachiller.	isDegreeOf	Student	Baccalaureate Degree
hasInformationAbout	Permite relacionar la información contenida en la ficha socioeconómica	informationIncludedIn	Socioeconomic Record	<ul style="list-style-type: none"> •FamilyHouse •FamilyAssets •FamilyGroup •Expense •Income
hasPart	Permite definir que algo es parte de otro.	isPartOf		
hasPrerequisite	Permite establecer los prerrequisitos de una asignatura.	prerequisiteFor	Requirement	Course
hasPreviousUniversity	Relaciona un estudiante con la universidad anterior en la que estudió.	isPreviousUniversityOf	Student	University
hasRegister	Relaciona un estudiante con su ficha socioeconómica.	isRegisterOf	Student	Socioeconomic Record
hasResidencePlace	Permite establecer el lugar de residencia de una persona.	isResidencePlaceOf	Person	Administrative Area
hasTimeSection	Determina la sección de una oferta académica.	isTimeSectionOf	AcademicOffer	TimeSection
isEquivalentTo	Permite establecer la equivalencia de una asignatura.	hasEquivalent	Course	Subject Collection
isFamilyHead	Relaciona un jefe de familia con su grupo familiar.	hasFamilyHead	FamilyHead	FamilyGroup



Propiedad	Descripción	Inversa	Dominio	Rango
isMain AuthorityOf	Permite establecer las autoridades principales de la universidad.	hasMain AuthorityOf	Authority	University
isMax AuthorityOf	Permite establecer la autoridad máxima de la universidad.	hasMax Authority	Committee	University
locatedIn	Relaciona un objeto con el lugar en cual se localiza.	locationOf	<ul style="list-style-type: none"> •Building •AcademicOffer •Room •Faculty 	<ul style="list-style-type: none"> •Geographic Coordinates •Building •Room
offers	Permite indicar que un objeto es ofertado por otro.	<ul style="list-style-type: none"> •offeredBy •offeredDuring 	<ul style="list-style-type: none"> •AcademicOffer •Faculty •Syllabus 	<ul style="list-style-type: none"> •Course •School •Semester
placeOfBirth	Relaciona una persona con su lugar de nacimiento.		Person	Administrative Area
relatedTo	Relaciona una organización con otra.	relatedBy	Organization	Organization
requires	Relaciona una asignatura con sus requisitos.	requiredBy	Course	Requirement
responsibleFor	Relaciona una oferta con el docente encargado.	responsibilityOf	Professor	AcademicOffer

Anexo F

Propiedades de datos definidas en el vocabulario UCuenca-Ontology

Propiedad	Descripción	Dominio	Rango
address	Dirección	FamilyHouse or Person	xsd:string
amount	Cantidad de dinero	Expense or Income	xsd:decimal
approvalStatus	Estado de aprobación en una asignatura (aprobado, reprobado o suspenso)	EducationalRecord	xsd:string
aprovalWay	Forma de aprobación en una asignatura	EducationalRecord	xsd:string
birthdate	Fecha de nacimiento	Person	xsd:date
bloodGroup	Grupo sanguíneo	Person	xsd:string
buildingMaterial	Material de construcción de la vivienda familiar	FamilyHouse	xsd:string
cableTVOrSimilar	Define si la familia dispone o de servicio de TV cable o similar	FamilyAssets	xsd:boolean
capacidad	Número de personas para las cuales es adecuado un espacio	Room	xsd:integer
cellphoneNumber	Número de celular	Person	xsd:string
courseClassHours	Número de horas de clase asignadas a un docente	Professor	xsd:integer



Propiedad	Descripción	Dominio	Rango
courseCode	Identificador de una asignatura	Course	xsd:string
courseCredits Number	Número de créditos de una asignatura	Course	xsd:integer
courseTitle	Nombre de una asignatura	Course	xsd:string
description	Descripción		xsd:string
disabilityType	Tipo de discapacidad (visual, auditiva, física, etc.)	Person	xsd:string
day	día de la semana	ClassSchedule	xsd:string
endDate	Fecha final de un intervalo	TimeInterval	xsd:string
endTime	Hora final de un evento	AcademicOffer	xsd:string
enrollmentStatus	Estado de una matrícula (pre-matriculado, matriculado)	Enrollment	xsd:string
ethnicity	Etnia con la cual se identifica una persona	Person	xsd:string
familiarRole	Rol familiar que cumple el jefe de familia	HeadOfFamily	xsd:string
familyHomeType	Tipo de vivienda familiar	FamilyHouse	xsd:string
fechaEgreso	Fecha de egreso de un estudiante	GraduationCertificate	xsd:date
finalGrade	Nota final obtenida en una asignatura	EducationalRecord	xsd:decimal
firstName	Nombre de una persona	Person	xsd:string
floorCount	Número de pisos de un edificio	Building	xsd:integer
floorNumber	Número de piso en el cual se ubica una habitación	Room	xsd:integer
gender	Género (masculino, femenino)	Person	xsd:string
graduationDate	Fecha de graduación	GraduationCertificate	xsd:date
graduationGrade	Nota final de graduación	GraduationCertificate	xsd:decimal
groupName	Identificador de un grupo ofertado para una asignatura	AcademicOffer	xsd:string



Propiedad	Descripción	Dominio	Rango
hasDisabilityCard	Tiene carné de discapacidad	Person	xsd:boolean
hasInsitutional Email	Dirección de correo electrónico institucional	Person	xsd:string
hasMaxCapacity	Número máximo de estudiantes en una clase ofertada	AcademicOffer	xsd:integer
hasMinCapacity	Número mínimo de estudiantes en una clase ofertada	AcademicOffer	xsd:integer
hasPersonalEmail	Dirección de correo electrónico personal	Person	xsd:string
hasValue	Valor de un criterio de evaluación	EvaluationCriteria	xsd:decimal
highSchool Graduate Year	Año de graduación como bachiller de un estudiante	Student	time:generalYear
hoursDuration	Duración en horas	Curriculum	xsd:decimal
identifier	Número de cedula o pasaporte que se utiliza como identificación.	Person	rdfs:Literal
isActive	Permite establecer si un objeto se encuentra activo o no.	Curriculum or Semester	xsd:boolean
lastName	Apellido de una persona	Person	xsd:string
laterality	Lateralidad (diestro, zurdo)	Person	xsd:string
latitude	Latitud geográfica	GeographicCoordinates	xsd:decimal
longitude	Longitud geográfica	GeographicCoordinates	xsd:decimal
maritalStatus	Estado civil (soltero, casado, viudo, etc.)	Person	xsd:string
membersUnder6 Years	Número de miembros del grupo familiar menores de 6 años	FamilyGroup	xsd:integer
monthlyPayment	Pago mensual por arriendo o hipoteca de la vivienda familiar	FamilyHouse	xsd:decimal
name	Nombre cualquiera		rdfs:Literal



Propiedad	Descripción	Dominio	Rango
nationalityES	Nacionalidad	Person	Country xsd:string
nonProductive Properties	Número de propiedades no productivas que tiene la familia	FamilyAssets	xsd:integer
classroomsNumber	Número de aulas que tiene un edificio	Building	xsd:integer
coursesNumber	Número de asignaturas que contiene una malla curricular	CurriculumMap	xsd:integer
facultiesNumber	Número de facultades que tiene una universidad	University	xsd:integer
numberOf Members	Número de miembros que conforman el grupo familiar	FamilyGroup	xsd:integer
disability Percentege	Porcentaje de discapacidad definido en el carné	Person	xsd:decimal
phoneNumber	Número telefónico	Person or Family-House	xsd:string
possessionType	Tipo de posesión de la vivienda familiar	FamilyHouse	xsd:string
productive Properties	Número de propiedades productivas que posee una familia	FamilyAssets	xsd:integer
professionOr Occupation	Profesión u ocupación del jefe de familia	HeadOfFamily	rdfs:Literal
refrendation Number	Número de refrendación del acta de grado	GraduationCertificate	xsd:string
residenceArea	Área donde se ubica la vivienda familiar (urbana, rural)	FamilyHouse	xsd:string
startDate	Fecha inicial de un intervalo	TimeInterval	xsd:string
startTime	Hora inicial de un evento	AcademicOffer	xsd:string
studentsNumber	Número de miembros del grupo familiar que estudian en niveles diferentes al superior	FamilyGroup	xsd:integer



Propiedad	Descripción	Dominio	Rango
studienIn AnotherCareer	Indica si un estudiante cursó una carrera diferente	Student	xsd:boolean
studyStartDate	Fecha de inicio de estudios	GraduationCertificate	xsd:date
surfaceArea	Tamaño de un aula de clases	Room	xsd:decimal
telephoneLines	Número de líneas telefónicas que posee una familia	FamilyAssets	xsd:integer
timeRepeat Course	Número de veces que un estudiante cursa una asignatura	EducationalRecord	xsd:integer
title	Título de un documento	Document	rdfs:Literal
titlingWorkGrade	Nota del trabajo de titulación	TitlingWork	xsd:decimal
totalNumberOf Levels	Número total de niveles que tiene una carrera	Curriculum	xsd:integer
universityServer Dependent	Indica si el estudiante depende de un servidor de la universidad	FamilyGroup	xsd:boolean
uri	URI de un trabajo de titulación	TitlingWork	xsd:anyURI
vacational Properties	Número de propiedades vacacionales que posee una familia	FamilyAssets	xsd:integer
vehicleAppraisal	Avalúo total de los vehículos que posee una familia	FamilyAssets	xsd:decimal
vehiclesNumber	Número de vehículos que posee una familia	FamilyAssets	xsd:integer

Anexo G

Descripción de clases definidas en la red ontológica UCuenca-Ontology

uri	http://ucuenca.edu.ec/ontology/AcademicOffer
rdf:Class	uco:AcademicOffer
rdfs:label	Oferta Académica
rdfs:Comment	Se entiende por oferta académica la composición de las asignaturas que la universidad está en disposición de impartir como docencia.

uri	http://ucuenca.edu.ec/ontology/ClassSchedule
rdf:Class	uco:ClassSchedule
rdfs:label	Horario de Clases
rdfs:Comment	Período temporal en el cual se lleva a cabo una clase. Especifica hora de inicio, hora de fin y lugar.

uri	http://ucuenca.edu.ec/ontology/CurriculumMap
rdf:Class	uco:CurriculumMap
rdfs:label	Malla Curricular
rdfs:Comment	Es el instrumento que contiene la organización de todas las asignaturas dentro de una carrera universitaria, con sus respectivos prerrequisitos y correquisitos.



uri	http://ucuenca.edu.ec/ontology/UniversityDegree
rdf:Class	uco:UniversityDegree
rdfs:label	Título Universitario
rdfs:subClassOf	hero:Degree
rdfs:Comment	Titulación universitaria, es una distinción otorgada por una universidad, generalmente después de la terminación exitosa de algún programa de estudios.

uri	http://ucuenca.edu.ec/ontology/TitlingWork
rdf:Class	uco:TitlingWork
rdfs:label	TrabajoDeTitulacion
rdfs:subClassOf	bibo:Document
rdfs:Comment	Trabajo realizado por el estudiante para satisfacer los requerimientos necesarios para obtener un título universitario.

uri	http://ucuenca.edu.ec/ontology/EducationalRecord
rdf:Class	uco:EducationalRecord
rdfs:label	Registro Academico
rdfs:Comment	Registro que contiene la información (estado de aprobación, forma de aprobación, número de matrícula, etc) de las asignaturas en las cuales el estudiante se ha matriculado.

uri	http://ucuenca.edu.ec/ontology/FamilyAssets
rdf:Class	uco:FamilyAssets
rdfs:label	Bienes Familiares
rdfs:Comment	Conjunto de propiedades o riquezas que pertenecen a un grupo familiar específico.

uri	http://ucuenca.edu.ec/ontology/FamilyGroup
rdf:Class	uco:FamilyGroup
rdfs:label	Grupo Familiar
rdfs:Comment	Es un conjunto de personas unidas o no por vínculos de sangre que comparten los gastos necesarios para su sustento.



uri	http://ucuenca.edu.ec/ontology/GraduationCertificate
rdf:Class	uco:GraduationCertificate
rdfs:label	ActaDeGrado
rdfs:Comment	El acta de grado es el documento legal que ampara al estudiante que ha culminado una carrera universitaria y ha obtenido el título universitario respectivo.

uri	http://ucuenca.edu.ec/ontology/KnowledgeArea
rdf:Class	uco:KnowledgeArea
rdfs:label	AreaDeConocimiento
rdfs:Comment	Las áreas de conocimiento son ámbitos específicos de conocimiento, cada una con su naturaleza particular y sus diferentes métodos para adquirir conocimientos.

uri	http://ucuenca.edu.ec/ontology/TitlingMechanism
rdf:Class	uco:TitlingMechanism
rdfs:label	MecanismoDeTitulacion
rdfs:Comment	Modalidad definida que un estudiante egresado cumplir para obtener un título universitario.

uri	http://ucuenca.edu.ec/ontology/SocioeconomicRecord
rdf:Class	uco:SocioeconomicRecord
rdfs:label	RegistroSocioeconomico
rdfs:Comment	Registro solicitado al estudiante, el mismo contiene información sobre la situación socioeconómica de la familia del estudiante (vivienda, bienes, grupo familiar). El objetivo de registrar dicha información es realizar un cálculo diferenciado en aquellos casos en los que el estudiante debe cancelar valores a la universidad.

uri	http://ucuenca.edu.ec/ontology/StudyModality
rdf:Class	uco:StudyModality
rdfs:label	ModalidadDeEstudios
rdfs:Comment	Es la forma de enseñanza aplicada en los distintos programas existentes en una universidad. Pueden ser: Presencial, semipresencial, a distancia.



uri	http://ucuenca.edu.ec/ontology/SubjectCollection
rdf:Class	uco:SubjectCollection
rdfs:label	ConjuntoDeAsignaturas
rdfs:Comment	Permite representar equivalencias entre asignaturas, ya que una asignatura es equivalente a una o al conjunto de varias asignaturas.

uri	http://ucuenca.edu.ec/ontology/TimeSection
rdf:Class	uco:TimeSection
rdfs:label	Seccion
rdfs:Comment	Referencia a la sección de tiempo en la cual se dictan clases de una asignatura en particular, la cual puede ser matutina, vespertina o nocturna.

uri	http://ucuenca.edu.ec/ontology/TrainingAxis
rdf:Class	uco:TrainingAxis
rdfs:label	Eje de formacion
rdfs:Comment	Son una parte fundamental del diseño curricular, ya que refieren a una serie de actividades humanas que poseen sus propios núcleos temáticos.

Anexo H

SQL para extracción de datos

```
-- +++++ INP.FACULTAD +++++
-- +++++
with
facultad_carreras as (select distinct
trim(academico.facultad_id_haciaarriba(crr.dependencia_id)) as facultad_id,
REGEXP_SUBSTR(translate(trim(academico.facultad_haciaarriba(
crr.dependencia_id)), 'ÁÉÍÓÚÑ ' '()."-\/*/-+', 'AEIOUN__'), '[A-Z]*')
facultad_link,
trim(academico.facultad_haciaarriba(crr.dependencia_id)) as facultad,
trim(crr.id) carrera_id,
REGEXP_SUBSTR(translate(trim(crr.nombre), 'ÁÉÍÓÚÑ ' '()."-\/*/-+', 'AEIOUN__'),
'[A-Z]*') carrera_link,
trim(crr.nombre) carrera, trim(crr.arecon_id) arecon_id
from academico.sga_matriculas mtr
inner join academico.sga_carreras crr on crr.id = mtr.carrera_id
and crr.esobligatoria='N'
and crr.escomun='N'
and crr.espostgrado='N'
and crr.estaactivo= 'S'
)
-- +++++ INP.PERIODO ACADEMICO +++++
-- +++++
,oferta_espacio as (
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
|| 'DIA' || dia || 'HORA' || replace(replace(hora_inicio, ' ', ''), ':', 'MIN')
Id_Record, trim(espfis_id) Data ,tipesp_id tipo_espacio
from academico.sga_horarios_ofertas ho
```



```
inner join
adminuc.espacios_fisicos ef
on ho.espfis_id =ef.id
)

-- ++++++ INP.ESTUDIANTE ++++++
-- ++++++
,
estudiantes as
(select replace(PN.person_id,' ','') person_id, replace(person_id,' ','')
identificacion,trim(PN.ubicac_id) ubicac_id,trim(PN.apellidos) apellidos,
trim(PN.nombres) nombres,trim(PN.fecha_nacimiento) fecha_nacimiento,
trim(PN.tipo_sangre) tipo_sangre,
decode(trim(PN.genero),'M','HOMBRE','F','MUJER',NULL) genero,
decode(trim(PN.diestro),'D','Diestro','Z','ZURDO',NULL) lateralidad,
decode(trim(PN.estado_civil),'D','DIVORSIADO(A)','S','SOLTERO(A)','V',
'VIUDO(A)','C','CASADO(A)','U','UNION LIBRE',null) estado_civil
from ADMINUC.PERSONAS_NATURALES PN INNER JOIN
ACADEMICO.SGA_CARRERAS_ESTUDIANTES CE ON PN.PERSON_ID=CE.PERSONA_ID
),

trabajos_titulacion as (select
'PA' || va.perlec_id || 'M' || va.malla_id || 'P' || va.placar_id || 'C' ||
va.carrera_id || 'E' || replace(va.persona_id,' ','') acta_id,
trim(va.paraacademicogeneral_id) id_trabajo_titulacion,
trim(va.nombre_tesis) trabajo_titulacion,
trim(va.total) total,
trim(va.calificacion) calificacion,
trim(dga_link_tesis) link_tesis,
trim(va.tutor_id) tutor_id
from academico.vst_actas va
inner join academico.sga_datos_graduados dg
on va.carrera_id=dg.dag_carrera_id and dg.dag_persona_id=va.persona_id
order by paraacademicogeneral_id
)

,
ficha_socioeconomica as (
select
'PA' || perlec_id || 'E' || replace(identificacion,' ','') id_fs,
```



```
replace(identificacion,' ','') estudiante_id,
trim(perlec_id ) perlec_id ,
UPPER(decode(trim(etnia),'No declara',null,trim(etnia))) etnia,
trim(tipo_discapacidad) tipo_discapacidad,
trim(porcentaje) porcentaje,
decode(trim(tienecarnet),'S','SI',null) tienecarnet,
case when trim(otra_universidad) like '%UNIVERS%' then REGEXP_SUBSTR(
translate(trim(otra_universidad),'ÁÉÍÓÚÑ ' '()."-\\/*/+','AEIOUN_'),'[A-Z]*')
else null end universidad_id,
decode(trim(estudio_otra_carrera),'SI','SI',null) estudio_otra_carrera,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') ||
'VI' vivienda_id, -- ===== VIVIENDA =====
trim(materiales_vivienda) materiales_vivienda,
trim(tipo_vivienda) tipo_vivienda,
REGEXP_SUBSTR(translate(trim(pais_vivienda_familiar),
'ÁÉÍÓÚÑ ' '()."-\\/*/+','AEIOUN_'),'[A-Z]*') pais_id,
REGEXP_SUBSTR(translate(trim(provincia_vivienda_familiar),
'ÁÉÍÓÚÑ ' '()."-\\/*/+','AEIOUN_'),'[A-Z]*') provincia_id,
REGEXP_SUBSTR(translate(trim(trim(canton_vivienda_familiar)),
'ÁÉÍÓÚÑ ' '()."-\\/*/+','AEIOUN_'),'[A-Z]*') canton_id,
upper(trim(direccion_familiar)) direccion_vivienda,
decode(trim(mensual_pago_arriendo),'0',null,trim(mensual_pago_arriendo))
pago_mensual,
decode(trim(telefono_familiar),'',null,trim(telefono_familiar)) telefono_casa,
trim(tenencia_vivienda) tipo_posesion,
trim(zona_vivienda) zona_vivienda,

'PA' || perlec_id || 'E' || replace(identificacion,' ','') ||
'GF' grupo_fam_id, -- ===== GRUPO FAMILIAR =====
trim(num_integrantes) num_miembros,
trim(rol_jefe_familiar) rol_jefe_familiar,
trim(num_estudiantes_familia) num_estudiantes_familia,
trim(num_hijos_men6_estudiante) um_hijos_men6_estudiante,
trim(es_carga_personal) es_carga_personal,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') ||
'BF' bienes_fam_id, -- ===== BIENES FAMILIARES =====
trim(tiene_tv_cable) tiene_tv_cable,
trim(nro_propiedades_no_renteras) propiedades_no_productivas,
trim(nro_propiedades_renteras) propiedades_productivas,
trim(nro_propiedades_vacacionales) propiedades_vacaciones,
```

```
trim(num_lineas_telefonicas) lineas_telefonicas,
trim(nro_vehiculos) vehiculos,
trim(avaluo_acumulado_vehiculos) avaluo_acumulado_vehiculos,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'IN_SUELDO'
ingreso_id1, -- ===== INGRESO =====
'SUELDO' nombre_ing1,
trim(sueldos) valor_ing1,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'IN_RENTAS'
ingreso_id2,
'RENTAS' nombre_ing2,
trim(rentas) valor_ing2,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'IN_PENSION'
ingreso_id3,
'PENSIONES' nombre_ing3,
trim(pensiones) valor_ing3,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'IN_OTRO'
ingreso_id4,
'OTRO' nombre_ing4,
trim(otros_ingresos) valor_ing4,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'EG_APORTEIESS'
egreso_id1, -- ===== EGRESO =====
'APORTE_IESS' nombre_eg1,
trim(aportes_iess) valor_eg1,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'EG_IMP_RENTA'
egreso_id2,
'IMPUESTO_RENTA' nombre_eg2,
trim(impuesto_renta) valor_eg2,
'PA' || perlec_id || 'E' || replace(identificacion,' ','') || 'EG_IMP_PREDIAL'
egreso_id3,
'IMPUESTO_PREDIAL' nombre_eg3,
trim(impuesto_predial) valor_eg3,
replace(REGEXP_REPLACE(cedula_jefe_familiar,'[- ]',''),' ','')
jefe_familiar_id, -- ===== JEFE FAMILIAR =====
replace(REGEXP_REPLACE(cedula_jefe_familiar,'[- ]',''),' ','')
identificacion_jefe,
upper(trim(apellidos_jefe_familiar)) apellidos,
upper(trim(nombres_jefe_familiar)) nombres,
trim(fecha_nacimiento_jefe_familiar) fecha_nacimiento,
upper(trim(rol_jefe_familiar)) rol,
upper(trim(ocupacion_jefe_familiar)) ocupacion
```



```
from GSOCIOECONOMICA.MV_FICHASSE
where perlec_id>114
),
estudios_secundarios as(select trim(replace(person_id,' ','')) Id_Record,
trim(tiptit_id) titulo_id, trim(univer_id) colegio_id, trim(anio_graduacion)
anio_graduacion from
adminuc.titulos_personas tp
inner join
adminuc.instituciones_educativas ie
on tp.univer_id=ie.id and ie.nivel ='SEC')

-- ++++++ INP.UBICACION ++++++
-- ++++++
, eft as (
select ef.id, ef.nombre,to_char(ef.bloque_id) bloque_id,to_char(ef.area) area,
to_char(ef.dependencia_id) dependencia_id,to_char(capacidad) capacidad,
to_char(piso) piso, tiposp_id tipo_espacio
from adminuc.espacios_fisicos ef
)
,
coordenadas as (
select 'COORDENADA0' id,0 dependencia_id, '-2.901086' latitud, '-79.010153'
longitud from dual union
select 'COORDENADA1' id,1 dependencia_id,'-2.910072' latitud, '-78.991249'
longitud from dual union
select 'COORDENADA2' id,2 dependencia_id,'-2.910671' latitud, '-78.991573'
longitud from dual union
select 'COORDENADA3' id,3 dependencia_id,'-2.901104' latitud, '-79.009729'
longitud from dual union
select 'COORDENADA4' id,4 dependencia_id,'-2.900630' latitud, '-79.009898'
longitud from dual union
select 'COORDENADA5' id,5 dependencia_id,'-2.901693' latitud, '-79.010071'
longitud from dual union
select 'COORDENADA6' id,6 dependencia_id,'-2.920205' latitud, '-79.025910'
longitud from dual union
select 'COORDENADA7' id,7 dependencia_id,'-2.899946' latitud, '-79.010185'
longitud from dual union
select 'COORDENADA8' id,8 dependencia_id,'-2.899767' latitud, '-79.009867'
longitud from dual union
select 'COORDENADA9' id,9 dependencia_id,'-2.900462' latitud, '-79.011389'
```



```
longitud from dual union
select 'COORDENADA10' id,10 dependencia_id,'-2.920137' latitud, '-79.025101'
longitud from dual union
select 'COORDENADA48' id,48 dependencia_id,'-2.902299' latitud, '-79.005331'
longitud from dual union
select 'COORDENADA73' id,73 dependencia_id,'-2.900877' latitud, '-79.009089'
longitud from dual union
select 'COORDENADA181' id,181 dependencia_id,'-2.900561' latitud, '-79.010453'
longitud from dual

), bloques_coordenadas as (
select distinct bloque_id,c.id coordenada_id, latitud,longitud
from adminuc.espacios_fisicos ef inner join coordenadas c
on ef.dependencia_id=c.dependencia_id
),ubicac as (
select
id,
replace(SYS_CONNECT_BY_PATH(trim(descripcion),'/'),'UBICACION GEOGRAFICA/', '')
ubicacion,
replace(SYS_CONNECT_BY_PATH(trim(id),'/'),'0/', '') ubicacion_id,
nacionalidad, level
from ADMINUC.UBICACIONES u , dual
START WITH id=0
CONNECT BY nocycle padre = prior id
),
lugares as (
select distinct
u.id,
REGEXP_SUBSTR(translate(trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 1)),
'ÁÉÍÓÚÑ ' '()."-\\*/-+', 'AEIOUN_'), '[A-Z]*') pais_id,
trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 1)) pais,
REGEXP_SUBSTR(translate(trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 2)),
'ÁÉÍÓÚÑ ' '()."-\\*/-+', 'AEIOUN_'), '[A-Z]*') provincia_id,
trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 2)) provincia,
REGEXP_SUBSTR(translate(trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 3)),
'ÁÉÍÓÚÑ ' '()."-\\*/-+', 'AEIOUN_'), '[A-Z]*') canton_id,
trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 3)) canton,
REGEXP_SUBSTR(translate(trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 4)),
'ÁÉÍÓÚÑ ' '()."-\\*/-+', 'AEIOUN_'), '[A-Z]*') parroquia_id,
trim(REGEXP_SUBSTR (ubicacion,'[/]+' ,1, 4)) parroquia,
```



```
nacionalidad
from
ubicac u
where id<>0
)
-- ++++++ INP.COLEGIOS ++++++
-- ++++++
,
colegios as(
select id, descripcion colegio, REGEXP_SUBSTR(translate(trim(descripcion),
'ÁÉÍÓÚÑ- ''.'()\/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') colegio_link
from adminuc.instituciones_educativas where nivel ='SEC'
),
-- ++++++ INP.TITULOS BACHILLER ++++++
-- ++++++
titulos_bachiller as(
select distinct id , nombre titulo, REGEXP_SUBSTR(translate(trim(nombre),
'ÁÉÍÓÚÑ- ''.'()\/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') titulo_link
from adminuc.titulos
where nivfor_id=1
),
-- ++++++ INP.TITULOS ++++++
-- ++++++
titulos_universitarios as(
select distinct id, REGEXP_SUBSTR(translate(trim(nombre), 'ÁÉÍÓÚÑ- ''.'()\/*/-+',
'AEIOUN__'), '[A-Z_0-9]*') titulo_link, trim(nombre) titulo,
trim(nombre_femenino) titulo_f
from adminuc.titulos ttl inner join ACADEMICO.SGA_TITULOS__CARRERAS ttlcrr
on ttlcrr.titulo_id=ttl.id and ttlcrr.envigencia = 'S'
),
-- ++++++ INP.areas_conocimiento ++++++
-- ++++++
areas_conocimiento as (
select distinct id, trim(nombre) area_conocimiento
, REGEXP_SUBSTR(translate(trim(nombre), 'ÁÉÍÓÚÑ- ''.'()\/*/-+', 'AEIOUN__'),
'[A-Z_0-9]*') id_link
from academico.sga_areas_conocimiento
)
-- ++++++ INP.DOCENTE ++++++
```



```
-- ++++++

,servidores as (select
replace(trim(pn.person_id),' ','') docente_id,
replace(trim(pn.person_id),' ','') identificacion,
trim(apellidos) apellidos,
trim(nombres) nombres,
ubicac_id,
trim(fecha_nacimiento) fecha_nacimiento,
trim(tipo_sangre) tipo_sangre,
decode(trim(genero),'M','HOMBRE','F','MUJER',NULL) genero,
decode(trim(diestro) ,'D','Diestro','Z','ZURDO',NULL) lateralidad,
decode(trim(PN.estado_civil),'D','DIVORSIADO(A)','S','SOLTERO(A)','V','VIUDO(A)',
,'C','CASADO(A)','U','UNION LIBRE',null) estado_civil
from ADMINUC.PERSONAS_NATURALES PN INNER JOIN GHUMANA.SERVIDORES@ASUCUENCA SE
ON PN.PERSON_ID=SE.PERSON_ID and SE.TIPO_SERVIDOR=3
)
-- ||||| INICIO DE CONSULTA |||||
select
replace(replace(Id_Record,chr(10),' '),chr(13),' ') Id_Record,
Field,
replace(replace(replace(Data,chr(10),' '),chr(13),' '),chr(11),' ') Data
from
(select
replace(Id_Record,' ','') Id_Record,
Field,
case when (length(Field) - length(replace(Field,'/',null))=2) then
replace(replace(replace(Data,chr(32),' '),chr(34),' '),chr(39),' ')
else replace(replace(Data,chr(34),' '),chr(39),' ') end Data
from (

-- %%%%%%%%%%% INICIO DE CONSULTA %%%%%%%%%%
-- %%%%%%%%%%%
-- %%%%%%%%%%% INICIO DE CONSULTA %%%%%%%%%%
-- %%%%%%%%%%%
-- %%%%%%%%%%%

-- =====
-- ===== COLEGIO =====
-- =====
```



162



```
select distinct facultad_id, facultad_link from facultad_carreras) fc
on dependencia_id=fc.facultad_id
inner join ADMINUC.BLOQUES b
on ef.bloque_id=b.id
union
select distinct facultad_link Id_Record, 'Facultad/esParteDe/Universidad' Field,
'UNIVERSIDAD_DE_CUENCA' Data
from facultad_carreras
union
select distinct facultad_link Id_Record, 'Facultad/relacionadoA/Departamento'
Field, 'vice_rectorado' Data
from facultad_carreras
-- =====
-- ===== CARRERAS =====
-- =====

union
select carrera_link Id_Record, Field, Data from
(select * from facultad_carreras fc inner join areas_conocimiento ac on
fc.arecon_id=ac.id)
unpivot(Data for Field in(
carrera as 'Carrera',
facultad_link as 'Carrera/ofrecidoPor/Facultad',
id_link as 'Carrera/esParteDe/AreaConocimiento'
))
union
select distinct carrera_link Id_Record, 'Carrera/entrega/TituloUniversitario'
Field, titulo_link Data
from ACADEMICO.SGA_TITULOS__CARRERAS tc inner join titulos_universitarios t
on tc.titulo_id=t.id inner join facultad_carreras fc on
tc.carrera_id= fc.carrera_id where tc.envigencia='S'

-- =====
-- ===== AREA CONOCIMIENTO =====
-- =====

union
select distinct id_link Id_Record, 'AreaConocimiento' Field, area_conocimiento
Data
from areas_conocimiento
```




```
-- =====
-- ===== TITULO UNIVERSITARIO =====
-- =====

UNION
select
titulo_link Id_Record, Field, Data
from
titulos_universitarios
unpivot (
Data for Field in
(titulo as 'TituloUniversitario',
titulo_f as 'TituloUniversitario/nombreFemenino'
))

-- =====
-- ===== PLAN CARRERA =====
-- =====

UNION
-- Se concatena placar, carrera para identificar de manera unica
--al plan de carrera
select distinct 'P' || id || 'C' || carrera_id Id_Record,
'PlanCarrera' Field, trim(nombre) Data
from ACADEMICO.SGA_PLANES_CARRERAS
union
select distinct 'P' || id || 'C' || carrera_id Id_Record,
'PlanCarrera/duracion' Field, trim(duracion) Data
from ACADEMICO.SGA_PLANES_CARRERAS
union
select distinct 'P' || id || 'C' || pc.carrera_id Id_Record,
'PlanCarrera/esParteDe/Carrera' Field, carrera_link Data
from ACADEMICO.SGA_PLANES_CARRERAS pc inner join facultad_carreras fc
on pc.carrera_id=fc.carrera_id
union
select distinct 'P' || id || 'C' || carrera_id Id_Record,
'PlanCarrera/enVigencia' Field, 'TRUE' Data
from ACADEMICO.SGA_PLANES_CARRERAS where envigencia='S'
union
select distinct 'P' || placar_id || 'C' || carrera_id Id_Record,
'PlanCarrera/tiene/ModalidadEstudio', REGEXP_SUBSTR(translate(trim(m.nombre),
'ÁÉÍÓÚÑ- ' ' ' .()\"/\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Data
```



```
from academico.sga_planes__modalidades pm
inner join academico.sga_modalidades m on pm.modalidad_id=m.id
-- =====
-- ===== MODALIDAD ESTUDIO =====
-- =====
union
select distinct REGEXP_SUBSTR(translate(trim(nombre),
'ÁÉÍÓÚÑ- ''.( )\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Id_Record,
'ModalidadEstudio', trim(nombre) Data from academico.sga_modalidades
-- =====
-- ===== MALLA =====
-- =====
UNION
-- Se concatena malla, placar, carrera para identificar de manera unica
--a la malla
select distinct 'M' || id || 'P' || placar_id || 'C' || carrera_id Id_Record,
'Malla' Field, trim(nombre) Data
from ACADEMICO.SGA_MALLAS
union
select distinct 'M' || id || 'P' || placar_id || 'C' || carrera_id Id_Record,
'Malla/tienNroAsignaturas' Field, trim(total_asignaturas) Data
from ACADEMICO.SGA_MALLAS
union
select distinct 'M' || id || 'P' || placar_id || 'C' || carrera_id Id_Record,
'Malla/esParteDe/PlanCarrera' Field, 'P' || placar_id || 'C' || carrera_id Data
from ACADEMICO.SGA_MALLAS

-- =====
-- ===== ASIGNATURA =====
-- =====
union
select distinct trim(id) Id_Record, 'Asignatura' Field, trim(nombre) Data
from
ACADEMICO.SGA_ASIGNATURAS
where ambito='PRE'
UNION
select distinct trim(id) Id_Record, 'Asignatura/tieneCodigo' Field,
replace(codigo, ' ', '') Data
from
ACADEMICO.SGA_ASIGNATURAS
```



```
where ambito='PRE' and codigo is not null
union
select distinct trim(id) Id_Record,'Asignatura/tieneCreditos' Field,
trim(nro_creditos) Data
from
ACADEMICO.SGA_ASIGNATURAS
where ambito='PRE' and nro_creditos is not null
union
select distinct trim(sa.id) Id_Record,'Asignatura/esParteDe/EjeFormacion'
Field, REGEXP_SUBSTR(translate(trim(ef.nombre),
'ÁÉÍÓÚÑ- ' '.()"\/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Data
from
ACADEMICO.SGA_ASIGNATURAS sa inner join
ACADEMICO.SGA_EJES_FORMACION ef on sa.ejefor_id=ef.id
where ambito='PRE' and ejefor_id is not null
-- =====
-- ===== NIVEL =====
-- =====
UNION
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'N'
|| trim(nivel_numero) Id_Record,'Nivel/tienexParte/Asignatura' Field,
trim(asignatura_id) Data
from ACADEMICO.SGA_ASIGNATURAS__MALLAS
union
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'N'
|| trim(nivel_numero) Id_Record,'Nivel' Field, trim(nivel_numero) Data
from ACADEMICO.SGA_ASIGNATURAS__MALLAS
union
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'N'
|| trim(nivel_numero) Id_Record,'Nivel/esParteDe/Malla' Field, 'M' || malla_id
|| 'P' || placar_id || 'C' || carrera_id Data
from ACADEMICO.SGA_ASIGNATURAS__MALLAS
-- =====
-- ===== EQUIVALENCIA =====
-- =====
UNION
select Id_Record, Field, trim(Data) from
(select distinct Id_Record,Field, Data
from
```



```
(
select trim(id) Id_Record, 'M' || id || 'P' || placar_id || 'C' ||
carrera_id malla, trim(asignatura_id) asignatura_id,
trim(asignatura1_id) asignatura1_id, trim(asignatura2_id) asignatura2_id,
trim(asignatura3_id) asignatura3_id, trim(asignatura4_id) asignatura4_id
from ACADEMICO.SGA_EQUIPARACIONES
)
unpivot
(Data for Field in(
malla as 'Equivalencia/esParteDe/Malla',
asignatura_id as 'Equivalencia/esEquivalenteDe/Asignatura',
asignatura1_id as 'Equivalencia/tienexParte/Asignatura',
asignatura2_id as 'Equivalencia/tienexParte/Asignatura',
asignatura3_id as 'Equivalencia/tienexParte/Asignatura',
asignatura4_id as 'Equivalencia/tienexParte/Asignatura'
))
) equiparaciones
where equiparaciones.Data is not null

-- =====
-- ===== REQUISITO =====
-- =====
UNION
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'A'
|| asignatura_id Id_Record, 'Requisito/tienexParte/Malla' Field, 'M' ||
prr_malla_id || 'P' || prr_placar_id || 'C' || prr_carrera_id Data
from ACADEMICO.SGA_PRERREQUISITOS
union
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'A'
|| asignatura_id Id_Record, 'Requisito/tienexPrerrequisito/Asignatura' Field,
trim(prr_asignatura_id) Data
from ACADEMICO.SGA_PRERREQUISITOS where escorrequisito='N'
union
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'A'
|| asignatura_id Id_Record, 'Requisito/tienexCorequisito/Asignatura' Field,
trim(prr_asignatura_id) Data
from ACADEMICO.SGA_PRERREQUISITOS where escorrequisito='S'
union
select distinct 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id || 'A'
|| asignatura_id Id_Record, 'Requisito/esRequisitoDe/Asignatura' Field,
```



```
trim(asignatura_id) Data
from ACADEMICO.SGA_PRERREQUISITOS
```

```
-- =====
-- ===== EJE FORMACION =====
-- =====
```

```
UNION
select distinct REGEXP_SUBSTR(translate(trim(nombre),
'ÁÉÍÓÚÑ- ''.( )\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Id_Record, 'EjeFormacion'
Field, trim(nombre) Data
from ACADEMICO.SGA_EJES_FORMACION
```

```
-- =====
-- ===== SILABO =====
-- =====
```

```
UNION
select Id_Record, Field, trim(Data) from (
select distinct trim(id_silabo) Id_Record , Field, Data
from
(
select trim(ss.id) id_silabo, trim(asignatura_id) asignatura_id,
trim(perlec_id) perlec_id, trim(cd.id) id_capitulo, trim(cd.contenido)
capitulo,
trim(ssu.id) id_subcapitulo, trim(ssu.tarea) subcapitulo
from SILABO.SGS_SILABO ss
inner join
SILABO.SGS_CONTENIDO_CABECERA sc
on ss.id=sc.silabo_id
inner join SILABO.SGS_CONTENIDO_DETALLE cd -- capitulo
on sc.id=cd.concab_id
inner join SILABO.SGS_SUBCAPITULOS ssu --subcapitulo
on cd.id=ssu.condet_id
)
)
unpivot (
Data for Field in(
asignatura_id as 'Silabo/esParteDe/Asignatura',
perlec_id as 'Silabo/esParteDe/PeriodoAcademico',
id_capitulo as 'Silabo/contiene/Capitulo'
)
)
```



```
) silabos2
where silabos2.Data is not null
union
select trim(silabo_id) Id_Record, 'Silabo/tienexCriterio/CriterioEvaluacion',
trim(id) Data
from SILABO.sgs_metodos_EVALUACION
```

```
-- =====
-- ===== CRITERIO EVALUACION =====
-- =====
```

```
union
select Id_Record,Field, Data
from
(
select trim(me.id) Id_Record, trim(peso) peso, trim(m.nombre) nombre
from SILABO.sgs_metodos_EVALUACION me
inner join
SILABO.sgs_metodos m
on me.metodos_id=m.ID
)
unpivot(
Data for Field in(
nombre as 'CriterioEvaluacion',
peso as 'CriterioEvaluacion/tienexPeso'
)
)
```

```
-- =====
-- ===== CAPITULO =====
-- =====
```

```
union
select distinct trim(id) Id_Record, 'Capitulo' Field,trim(contenido) Data
from SILABO.SGS_CONTENIDO_DETALLE
/*union
select distinct trim(condet_id) Id_Record, 'Capitulo/tienexParte/Subcapitulo'
Field, trim(id) Data from SILABO.SGS_SUBCAPITULOS
*/
```

```
-- =====
-- ===== SUBCAPITULO =====
```



```
-- =====
/*union
select distinct trim(id) Id_Record, 'Subcapitulo' Field, trim(tarea) Data from
SILABO.SGS_SUBCAPITULOS
*/
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%      INP. PERIODO ACADEMICO      %%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
UNION
-- =====
-- ===== PERIODO ACADEMICO =====
-- =====
select distinct trim(id) Id_Record,Field, trim(Data) Data
from ADMINUC.periodos_lectivos
unpivot (Data for Field in(
fechainicio as 'PeriodoAcademico/inicio',
fechafin as 'PeriodoAcademico/fin'
))

-- =====
-- ===== OFERTA ACADEMICA =====
-- =====
UNION
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
Id_Record,
'OfertaAcademica/ofertadoX/PeriodoAcademico' Field, trim(perlec_id) Data
from ACADEMICO.sga_oferta_asignaturas
union
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
Id_Record,
'OfertaAcademica/oferta/Asignatura' Field, trim(asignatura_id) Data
from ACADEMICO.sga_oferta_asignaturas
union
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
```



```
Id_Record,
'OfertaAcademica/tema' Field, trim(nombre_tema) Data
from ACADEMICO.sga_oferta_asignaturas where nombre_tema is not null
union
select distinct
'PA' || oa.perlec_id || 'C' || oa.carrera_id || 'A' || oa.asignatura_id || 'G'
|| oa.grupo Id_Record,
'OfertaAcademica/tienexParte/Seccion' Field, REGEXP_SUBSTR(translate(
trim(s.nombre),'ÁÉÍÓÚÑ-''.'()\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Data
from ACADEMICO.sga_oferta_asignaturas oa
inner join
ACADEMICO.SGA_SECCIONES s
on oa.seccion_id=s.id
where seccion_id is not null
union
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
Id_Record,
'OfertaAcademica/ofertaEnGrupo' Field, trim(grupo) Data
from ACADEMICO.sga_oferta_asignaturas where grupo is not null
union
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
Id_Record,
'OfertaAcademica/tieneCupoMinimo' Field, trim(cupo_min) Data
from ACADEMICO.sga_oferta_asignaturas where cupo_min is not null
union
select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
Id_Record,
'OfertaAcademica/tieneCupoMaximo' Field, trim(cupo_max) Data
from ACADEMICO.sga_oferta_asignaturas where cupo_max is not null

union

select distinct
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G'
|| grupo Id_Record,
'OfertaAcademica/tienexEncargado/Docente' Field, trim(docente_id) Data
from ACADEMICO.sga_oferta_docentes where esprincipal='S'
```




```
union
SELECT distinct 'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id
|| 'G' || grupo Id_Record,
'OfertaAcademica/tienexParte/Horario' Field,
'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G' || grupo
|| 'DIA' || dia || 'HORA' || replace(replace(hora_inicio,' ',''),':','MIN') Data
from ACADEMICO.SGA_HORARIOS_OFERTAS
union
SELECT distinct 'PA' || perlec_id || 'C' || oa.carrera_id || 'A'
|| asignatura_id || 'G' || grupo Id_Record,
'OfertaAcademica/ofertadoPor/Carrera' Field, carrera_link Data
from ACADEMICO.sga_oferta_asignaturas oa
inner join
facultad_carreras fc
on oa.carrera_id=fc.carrera_id
where PERLEC_ID>108
union
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%      INP. HORARIO      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
SELECT distinct Id_Record, Field, Data FROM
(select 'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id || 'G'
|| grupo || 'DIA' || dia || 'HORA' || replace(replace(hora_inicio,' ',''),':',
'MIN') Id_Record, hora_inicio,hora_fin,
decode(dia,1,'LUNES',2,'MARTES',3,'MIERCOLES',4,'JUEVES',5,'VIERNES',6
,'SABADO',7,'DOMINGO',NULL) dia from
ACADEMICO.SGA_HORARIOS_OFERTAS)
UNPIVOT
(Data for Field in
(hora_inicio as 'Horario/horaInicio',
hora_fin as 'Horario/horaFin',
dia as 'Horario/dia'))
union
select distinct Id_Record, 'Horario/dictadoEn/Laboratorio' Field, Data
from oferta_espacio
where tipo_espacio=1
union
select distinct Id_Record, 'Horario/dictadoEn/Auditorio' Field, Data
```



```
from oferta_espacio where tipo_espacio=2
union
select distinct Id_Record, 'Horario/dictadoEn/Taller' Field, Data
from oferta_espacio where tipo_espacio=3
union
select distinct Id_Record, 'Horario/dictadoEn/Sala' Field, Data
from oferta_espacio where tipo_espacio=4
union
select distinct Id_Record, 'Horario/dictadoEn/Aula' Field, Data
from oferta_espacio where tipo_espacio=5
union
select distinct Id_Record, 'Horario/dictadoEn/Oficina' Field, Data
from oferta_espacio where tipo_espacio=6

-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%      INP. ESTUDIANTE      %%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
UNION
-- =====
-- ===== ESTUDIANTES =====
-- =====

select distinct trim(person_id) Id_Record, Field, Data from estudiantes
unpivot (
Data FOR Field IN (
identificacion as 'Estudiante',
apellidos as 'Estudiante/tienexApellidos',
nombres as 'Estudiante/tienexNombres',
fecha_nacimiento as 'Estudiante/tienexFechaNacimiento',
tipo_sangre as 'Estudiante/esDeGrupoSanguineo' ,
genero as 'Estudiante/tienexGenero',
lateralidad as 'Estudiante/tienexLateralidad',
estado_civil as 'Estudiante/tienexEstadoCivil'
))
WHERE Data is not null
union
SELECT replace(trim(id),' ','') Id_Record, 'Estudiante/tienexDireccion' Field,
upper(trim(direccion)) Data FROM
```



```
ADMINUC.PERSONAS PN
inner join
estudiantes e
on e.identificacion=PN.id
union
select trim(Id_Record) Id_Record, Field,Data
from
(
select personas.* from
(select replace(p.person_id,' ','') Id_Record,
'Estudiante/emailPersonal' Field, lower(trim(p.email_personal)) Data
from ADMINUC.PERSONAS_NATURALES p
where instr(email_personal,'@',3)>2
union
select replace(p.ID,' ','') Id_Record,
'Estudiante/emailInstitucional' Field, lower(trim(email)) Data
from ADMINUC.PERSONAS p where instr(email,'@',3)>2
union
select replace(person_id,' ','') Id_Record,case
when tiptel_id=1 then 'Estudiante/telefonoConvecional'
else 'Estudiante/telefonoCelular' end Field , trim(numero_telefono) Data
from adminuc.telefonos_personas where tiptel_id=2 or tiptel_id=1
) personas
inner join estudiantes e
on personas.Id_Record=e.person_id
)
union
select distinct replace(persona_id,' ','') Id_Record,
'Estudiante/ingresoPor/Admision' Field, trim(plan_contingencia) Data
from ACADEMICO.SGA_CARRERAS__ESTUDIANTES CE
union
select distinct
trim(persona_id) Id_Record, 'Estudiante/tienexMatricula/Matricula' Field,
'PA' || perlec_id || 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id
|| 'E' || replace(persona_id,' ','') Data
from ACADEMICO.sga_matriculas

union

select distinct
```



```
trim(P.id) Id_Record,
case
when provincia_id is null then 'Estudiante/tienexPais/Pais'
when canton_id is null then 'Estudiante/tienexProvincia/Provincia'
when parroquia is null then 'Estudiante/tienexCanton/Canton'
else 'Estudiante/tienexParroquia/Parroquia' end as Field,

case
when provincia_id is null then pais_id
when canton_id is null then provincia_id
when parroquia is null then canton_id
else parroquia_id end as Data

from
lugares l
inner join
ADMINUC.PERSONAS P
on l.id=P.ubicac_id
INNER JOIN  ACADEMICO.SGA_CARRERAS__ESTUDIANTES CE
ON P.ID=CE.PERSONA_ID

union
select
Id_Record, Field, Data
from
(select Id_Record, titulo_link, colegio_link, anio_graduacion
from estudios_secundarios es
inner join
titulos_bachiller tb
on es.titulo_id=tb.id
inner join
colegios c
on es.colegio_id=c.id
)
unpivot(
Data for Field in(
titulo_link as 'Estudiante/tienexTitulo/TituloBachiller',
colegio_link as 'Estudiante/tienexColegio/Colegio',
anio_graduacion as 'Estudiante/tienexAnioGraduacion'
)
```



```
)
union
select distinct replace(estudiante_id,' ','') Id_Record, Field,Data
from (select distinct * from ficha_socioeconomica)
unpivot (Data for Field in(
etnia as 'Estudiante/tieneEtnia',
tipo_discapacidad as 'Estudiante/tieneDiscapacidad',
porcentaje as 'Estudiante/tiene/porcentajeDiscapacidad',
tienecarnet as 'Estudiante/tieneCarnet',
universidad_id as 'Estudiante/vieneDe/Universidad',
estudio_otra_carrera as 'Estudiante/otraCarrera'))
union
select distinct replace(ce.persona_id,' ','') Id_Record,
'Estudiante/esParteDe/Carrera' Field, fc.carrera_link Data
from ACADEMICO.SGA_CARRERAS__ESTUDIANTES ce
inner join
facultad_carreras fc
on ce.carrera_id=fc.carrera_id
-- =====
-- ===== ADMISION =====
-- =====
UNION

select valor Id_Record,'Admision' Field, trim(nombre) Data
from ACADEMICO.SGA_TIPOS_INSCRIPCIONES

-- =====
-- ===== MATRICULA =====
-- =====
UNION
select distinct
'PA' || perlec_id || 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id
|| 'E' || replace(persona_id,' ','') Id_Record,
'Matricula/tienexEstado' Field, decode(trim(estado_matricula),'I','INCOMPLETO',
'G','GRABADO','R','REGISTRADO','P','PREMATRICULADO','M','MATRICULADO') Data
from ACADEMICO.sga_matriculas where perlec_id>108
union
select distinct
'PA' || perlec_id || 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id
```



```
|| 'E' || replace(persona_id,' ','') Id_Record,
'Matricula/tienexMalla/Malla' Field,
'M' || malla_id || 'P' || placar_id || 'C' || carrera_id Data
from ACADEMICO.sga_matriculas where perlec_id>108
/*
union
select distinct
'PA' || perlec_id || 'M' || malla_id || 'P' || placar_id || 'C' || carrera_id
|| 'E' || replace(persona_id,' ','') Id_Record,
'Matricula/tienexParte/RegistroAcademico' Field,
'PA' || placar_id || 'C' || carrera_id || 'E' || persona_id Data
from ACADEMICO.sga_matriculas
*/
```

```
-- =====
-- ===== REGISTRO =====
-- =====
-- RELACIONA el registro academico con la matricula y la oferta academica
UNION
```

```
select * from
(
select distinct 'PA' || perlec_id || 'C' || carrera_id || 'A' || asignatura_id
|| 'G' || grupo || 'E' || replace(persona_id,' ','') Id_Record,
'PA' || perlec_id || 'C' || carrera_toma || 'A' || asignatura_id || 'G' ||
grupo oferta_academica,
to_char(nota_final) nota_final,to_char(veces_cursa) veces_cursa,
decode(trim(estado_aprobacion),'AP','APROBADO','SU','SUSPENSO','RE',
'REPROBADO','RF','REPROBADO-FALTAS','RS','REPROBADO-SANCION','SS',
'SUSPENSO-SANCION','RR','REPROBADO-REGLAMENTO CARRERA') estado_aprobacion
from ACADEMICO.sga_matriculas__asignaturas
where grupo<>0 and esta_anulada<>'S' and perlec_id>108
)
unpivot (
Data FOR Field IN (
nota_final as 'RegistroAcademico/tienexCalificacion',
estado_aprobacion as 'RegistroAcademico/tieneEstadoAprobacion',
veces_cursa as 'RegistroAcademico/vecesCursa',
oferta_academica as 'RegistroAcademico/registra/OfertaAcademica'
))
```



```
union
select distinct
'PA' || ma.perlec_id || 'C' || ma.carrera_id || 'A' || ma.asignatura_id || 'G'
|| ma.grupo || 'E' || replace(ma.persona_id, ' ', '') Id_Record,
'RegistroAcademico/esParteDe/Matricula' Field,
'PA' || m.perlec_id || 'M' || m.malla_id || 'P' || m.placar_id || 'C' ||
m.carrera_id || 'E' || replace(m.persona_id, ' ', '') Data
from ACADEMICO.sga_matriculas__asignaturas ma inner join
ACADEMICO.sga_matriculas m
on ma.perlec_id=m.perlec_id and ma.carrera_id=m.carrera_id and
ma.persona_id=m.persona_id and ma.grupo<>0 and ma.esta_anulada<>'S'
and ma.perlec_id>108
-- =====
-- ===== ACTA GRADO =====
-- =====
UNION
select distinct Id_Record, Field, Data
from
(
select 'PA' || perlec_id || 'M' || malla_id || 'P' || placar_id || 'C' ||
carrera_id || 'E' || replace(persona_id, ' ', '') Id_Record,
'M' || malla_id || 'P' || placar_id || 'C' || carrera_id malla_id,
replace(trim(ae.titulo_id), ' ', '') titulo_id,
trim(dag_persona_id) graduado_id,
trim(DAG_NOTATRABAJOTITULACION) nota_trabajotitulacion,
-- trim(dga_link_tesis) link_tesis,
trim(dag_fechainicioestudios) inicio_estudios,
trim(dag_fechaegresamiento) fecha_egreso,
trim(dag_fechaactagrado) fecha_actagrado,
trim(dag_notapromedioacumulado) nota_promedioacumulado,
trim(dga_num_refrendacion) num_refrendacion,
-- trim(ae.nombre_tesis) trabajo_titulacion,
REGEXP_SUBSTR(translate(trim(pa.nombre),
'ÁÉÍÓÚÑ- ' '().\"/\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') mecanismo_link,
trim(perlec_id) perlec_id
from
academico.sga_datos_graduados dg
inner join
academico.sga_acta_estudiantes ae
on dg.dag_numero_acta=ae.numero_acta and dg.dag_carrera_id=ae.carrera_id
```



```
and dg.dag_persona_id=ae.persona_id
inner join
PREREQUISITO.PARAACADEMICO pa
on pa.id=ae.paraacademico_id
)
unpivot (
Data for Field in(
graduado_id as 'ActaGrado/tienexGraduado/Estudiante',
malla_id as 'ActaGrado/tienexMalla/Malla',
titulo_id as 'ActaGrado/tienexTitulo/TituloUniversitario',
fecha_egreso as 'ActaGrado/tienexFechaEgreso',
fecha_actagrado as 'ActaGrado/tienexFechaActa',
inicio_estudios as 'ActaGrado/tieneFechaInicio',
nota_promedioacumulado as 'ActaGrado/tienexNotaAcumulado',
-- nota_trabajotitulacion as 'ActaGrado/tienexNotaTrabajo',
num_refrendacion as 'ActaGrado/tienexNumRefrendacion',
-- trabajo_titulacion as 'ActaGrado/tienexTitulo',
-- link_tesis as 'ActaGrado/tienexEnlace',
mecanismo_link as 'ActaGrado/tiene/MecanismoTitulacion',
perlec_id as 'ActaGrado/tienexPeriodo/PeriodoAcademico'
))
union
select distinct acta_id Id_Rescord, 'ActaGrado/tiene/TrabajoTitulacion' Field,
id_trabajo_titulacion Data
from trabajos_titulacion
-- =====
-- ===== TRABAJO TITULACION =====
-- =====
UNION
select distinct id_trabajo_titulacion Id_Record, Field, Data
from trabajos_titulacion
unpivot
(Data For Field in(
trabajo_titulacion as 'TrabajoTitulacion',
-- acta_id as 'TrabajoTitulacion/tienexActa/ActaGrado',
total as 'TrabajoTitulacion/tienexTotal',
calificacion as 'TrabajoTitulacion/tienexCalificacion',
link_tesis as 'TrabajoTitulacion/tienexEnlace',
tutor_id as 'TrabajoTitulacion/dirigidoPor/Tutor'
))
```




```
-- =====
-- ===== MECANISMO TITULACION =====
-- =====
UNION
select distinct REGEXP_SUBSTR(translate(trim(pa.nombre),
'ÁÉÍÓÚÑ- ' '().\"/*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Id_Record ,
'MecanismoTitulacion' Field, pa.nombre Data
from PREREQUISITO.PARAACADEMICO pa
inner join academico.sga_acta_estudiantes ae on pa.id=ae.paraacademico_id
-- =====
-- ===== FICHA SOCIOECONOMICA =====
-- =====
UNION
select id_fs Id_Record, Field, Data
from ficha_socioeconomica
unpivot(Data for Field in(
estudiante_id as 'FichaSocioeconomica/tiene/Estudiante',
vivienda_id as 'FichaSocioeconomica/tiene/Vivienda',
perlec_id as 'FichaSocioeconomica/tiene/PeriodoAcademico',
grupo_fam_id as 'FichaSocioeconomica/tiene/GrupoFamiliar',
bienes_fam_id as 'FichaSocioeconomica/tiene/BienesFamiliares',
ingreso_id1 as 'FichaSocioeconomica/tienexIngreso/Ingreso',
ingreso_id2 as 'FichaSocioeconomica/tienexIngreso/Ingreso',
ingreso_id3 as 'FichaSocioeconomica/tienexIngreso/Ingreso',
ingreso_id4 as 'FichaSocioeconomica/tienexIngreso/Ingreso',
egreso_id1 as 'FichaSocioeconomica/tienexIngreso/Egreso',
egreso_id2 as 'FichaSocioeconomica/tienexIngreso/Egreso',
egreso_id3 as 'FichaSocioeconomica/tienexIngreso/Egreso'
))
union
select vivienda_id Id_Record, Field, Data
from ficha_socioeconomica
unpivot(Data for Field in(
materiales_vivienda as 'Vivienda/tienexMateriales',
tipo_vivienda as 'Vivienda/tienexTipo',
direccion_vivienda as 'Vivienda/tienexDireccion',
pago_mensual as 'Vivienda/tienexPagoMensual',
telefono_casa as 'Vivienda/tienexTelefono',
```



```
tipo_posesion as 'Vivienda/tienexPosesion',
zona_vivienda as 'Vivienda/tienexZona'
))
union
select vivienda_id Id_Record,
case
when provincia_id is null then 'Vivienda/tienexPais/Pais'
when canton_id is null then 'Vivienda/tienexProvincia/Provincia'
else 'Vivienda/tienexCanton/Canton' end as Field,
case
when provincia_id is null then pais_id
when canton_id is null then provincia_id
else canton_id end as Field
from ficha_socioeconomica
union
select grupo_fam_id Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
num_miembros as 'GrupoFamiliar/tienexNumMiembros',
num_estudiantes_familia as 'GrupoFamiliar/tienexNumEstudiantes',
um_hijos_men6_estudiante as 'GrupoFamiliar/tienexHijos6',
es_carga_personal as 'GrupoFamiliar/tienexCargaFamiliar',
jefe_familiar_id as 'GrupoFamiliar/tienexJefe/JefeFamiliar'
))
union
select bienes_fam_id Id_Record, Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
tiene_tvcable as 'BienesFamiliares/tieneTVCable',
propiedades_no_productivas as 'BienesFamiliares/propiedadesNoProductivas',
propiedades_productivas as 'BienesFamiliares/propiedadesProductivas',
propiedades_vacaciones as 'BienesFamiliares/propiedadesVacaciones',
lineas_telefonicas as 'BienesFamiliares/tienexNumLineasTelefonicas',
vehiculos as 'BienesFamiliares/tienexNumVehiculos',
avaluo_acumulado_vehiculos as 'BienesFamiliares/tienexAvaluo'
))
union
select ingreso_id1 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
```

```
nombre_ing1 as 'Ingreso/tienexDescripcion',
valor_ing1 as 'Ingreso/tienexValor'
))
union
select ingreso_id2 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_ing2 as 'Ingreso/tienexDescripcion',
valor_ing2 as 'Ingreso/tienexValor'
))
union
select ingreso_id3 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_ing3 as 'Ingreso/tienexDescripcion',
valor_ing3 as 'Ingreso/tienexValor'
))
union
select ingreso_id4 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_ing4 as 'Ingreso/tienexDescripcion',
valor_ing4 as 'Ingreso/tienexValor'
))
union
select egreso_id1 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_eg1 as 'Egreso/tienexDescripcion',
valor_eg1 as 'Egreso/tienexValor'
))
union
select egreso_id2 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_eg2 as 'Egreso/tienexDescripcion',
valor_eg2 as 'Egreso/tienexValor'
))
union
```



```
select egreso_id3 Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
nombre_eg3 as 'Egreso/tienexDescripcion',
valor_eg3 as 'Egreso/tienexValor'
))

union

select jefe_familiar_id Id_Record,Field,Data
from ficha_socioeconomica
unpivot(Data for Field in(
identificacion_jefe as 'JefeFamiliar',
apellidos as 'JefeFamiliar/tienexApellidos',
nombres as 'JefeFamiliar/tienexNombres',
fecha_nacimiento as 'JefeFamiliar/tienexFechaNacimiento',
rol as 'JefeFamiliar/tienexRol',
ocupacion as 'JefeFamiliar/tienexOcupacion'
))

-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% INP. UBICACION %%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
UNION
-- =====
-- ===== PAIS =====
-- =====

select distinct pais_id Id_Record, Field, Data
from lugares
unpivot
(Data for Field in(
pais as 'Pais',
provincia_id as 'Pais/tienexProvincia/Provincia',
nacionalidad as 'Pais/tienexNacionalidad'
)
)
where Data is not null
```



```
-- =====  
-- ===== PROVINCIA =====  
-- =====
```

```
union  
select distinct provincia_id Id_Record, Field, Data  
from lugares  
unpivot  
(Data for Field in(  
provincia as 'Provincia',  
canton_id as 'Provincia/tienexCanton/Canton'  
)  
)  
where Data is not null
```

```
-- =====  
-- ===== CANTON =====  
-- =====
```

```
union  
select distinct canton_id Id_Record, Field, Data  
from lugares  
unpivot  
(Data for Field in(  
canton as 'Canton',  
parroquia_id as 'Canton/tienexParroquia/Parroquia'  
)  
)  
where Data is not null
```

```
union  
-- =====  
-- ===== PARROQUIA =====  
-- =====
```

```
select parroquia_id Id_Record, 'Parroquia' Field, parroquia Data  
from lugares  
where parroquia_id is not null
```

```
-- =====  
-- ===== SECCION =====  
-- =====
```



```
UNION
SELECT REGEXP_SUBSTR(translate(trim(s.nombre), 'ÁÉÍÓÚÑ- ' '().\"/*/-+',
'AEIOUN__'), '[A-Z_0-9]*') Id_Record,
'Seccion' Field, nombre Data FROM ACADEMICO.SGA_SECCIONES s
```

```
-- =====
-- ===== ESPACIOS FISICOS =====
-- =====
```

```
union
select trim(id) Id_Record, Field, Data
from eft
unpivot(
Data for Field in(
nombre as 'Laboratorio',
capacidad as 'Laboratorio/tienexCapacidad',
piso as 'Laboratorio/piso',
area as 'Laboratorio/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Laboratorio/tienexBloque/Bloque'
)
)
where tipo_espacio=1
union
select trim(id) Id_Record, Field, Data
from eft
unpivot(
Data for Field in(
nombre as 'Auditorio',
capacidad as 'Auditorio/tienexCapacidad',
piso as 'Auditorio/piso',
area as 'Auditorio/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Auditorio/tienexBloque/Bloque'
)
)
where tipo_espacio=2
union
select trim(id) Id_Record, Field, Data
from eft
unpivot(
```



```
Data for Field in(
nombre as 'Taller',
capacidad as 'Taller/tienexCapacidad',
piso as 'Taller/piso',
area as 'Taller/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Taller/tienexBloque/Bloque'
)
)
where tipo_espacio=3
union
select trim(id) Id_Record, Field, Data
from eft
unpivot(
Data for Field in(
nombre as 'Sala',
capacidad as 'Sala/tienexCapacidad',
piso as 'Sala/piso',
area as 'Sala/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Sala/tienexBloque/Bloque'
)
)
where tipo_espacio=4
union
select trim(id) Id_Record, Field, Data
from eft
unpivot(
Data for Field in(
nombre as 'Aula',
capacidad as 'Aula/tienexCapacidad',
piso as 'Aula/piso',
area as 'Aula/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Aula/tienexBloque/Bloque'
)
)
where tipo_espacio=5
union
select trim(id) Id_Record, Field, Data
```

```
from eft
unpivot(
Data for Field in(
nombre as 'Oficina',
capacidad as 'Oficina/tienexCapacidad',
piso as 'Oficina/piso',
area as 'Oficina/area',
-- dependencia_id as 'EspacioFisico/facultad',
bloque_id as 'Oficina/tienexBloque/Bloque'
)
)
where tipo_espacio=6
-- =====
-- ===== BLOQUE =====
-- =====

union
select bloque_link Id_Record, Field, Data
from
(
select id,REGEXP_SUBSTR(translate(trim(nombre),
'ÁÉÍÓÚÑ- ''.'\"\\*/-+', 'AEIOUN__'), '[A-Z_0-9]*') bloque_link, nombre,
to_char(numero_aulas) nro_aulas, to_char(numero_pisos) nro_pisos
from ADMINUC.BLOQUES b
)
unpivot(
Data for Field in(
nombre as 'Bloque',
nro_aulas as 'Bloque/nroAulas',
nro_pisos as 'Bloque/nroPisos'
)
)
union

select REGEXP_SUBSTR(translate(trim(b.nombre),
'ÁÉÍÓÚÑ- ''.'\"\\*/-+', 'AEIOUN__'), '[A-Z_0-9]*') Id_Record,
'Bloque/tieneCoordenada/Coordenada' Field, bc.coordenada_id Data
from bloques_coordenadas bc inner join ADMINUC.BLOQUES b on bc.bloque_id=b.id

--unpivot
---( Data for Field in (
```




```
--      latitud as 'edificio/latitud',
--      longitud as 'edificio/longitud'
---))

-- =====
-- ===== COORDENADAS =====
-- =====

UNION
select trim(id) Id_Record, Field, Data
from coordenadas
unpivot
( Data for Field in (
latitud as 'Coordenada/tieneLatitud',
longitud as 'Coordenada/tieneLongitud'
)
)

-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%      INP. DOCENTE      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-- %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

UNION
select docente_id Id_Record, Field, Data from servidores
unpivot(Data for Field in(
identificacion as 'Docente',
apellidos as 'Docente/tienexApellidos',
nombres as 'Docente/tienexNombres',
fecha_nacimiento as 'Docente/tienexFechaNacimiento',
tipo_sangre as 'Docente/tienexGrupoSanguineo',
genero as 'Docente/tienexGenero',
lateralidad as 'Docente/tienexLateralidad'
))
union
select * from
(
select personas.* from
(select p.person_id Id_Record, 'Docente/emailPersonal' Field,
lower(trim(p.email_personal)) Data
from ADMINUC.PERSONAS_NATURALES p
where instr(email_personal, '@', 3) > 2
```



```
union
select p.ID Id_Record,'Docente/emailInstitucional' Field,
lower(trim(email)) Data from ADMINUC.PERSONAS p
where instr(email,'@',3)>2
union
select person_id Id_Record,case when tiptel_id=1
then 'Docente/telefonoConveccional' else 'Docente/telefonoCelular'
end Field , trim(numero_telefono) Data
from adminuc.telefonos_personas where tiptel_id=2 or tiptel_id=1
) personas
inner join servidores e
on personas.Id_Record=e.docente_id
)
union
select distinct
replace(trim(P.docente_id),' ','') Id_Record,
case
when provincia_id is null then 'Docente/tienexPais/Pais'
when canton_id is null then 'Docente/tienexProvincia/Provincia'
when parroquia_id is null then 'Docente/tienexCanton/Canton'
else 'Docente/tienexParroquia/Parroquia' end as Field,
case
when provincia_id is null then pais_id
when canton_id is null then provincia_id
when parroquia_id is null then canton_id
else parroquia_id end as Data

from
lugares l
inner join
servidores P
on l.id=P.ubicac_id

) where Data is not null)
```

Anexo I

Problemas y soluciones

Se documentó algunos problemas a los que se enfrentó y las solución propuesta durante el proceso de integración de datos. La finalidad este anexo es proveer soporte a los futuros proyectos que intenten replicar el proceso que se ha llevado acabo en este trabajo.

I.1. Extracción de datos

Para la extracción de datos, se sugiere tener en cuenta los caracteres especiales puesto que esto podría generar errores en las etapas posteiores.

I.2. Transformaciones

En esta sección se describe problemas relacionados a la herramienta PDI, y las transformaciones utilizadas para el proceso de generación del RDF.

I.2.1. Ampliación de memoria

La sobre carga de memoria es un problema recurrente al cual se ha enfrentado en este proyecto, esto debido a la gran catidad de datos con las que se trabaja, la solución a este problema es aumentar la cantidad de memoria asignada a PDI. Para hacerlo se deberá dirigir al directorio *dataintegration/spoon.sh* en Linux o *dataintegration/spoon.bat* en Windows, abrir el archivo con un editor de texto y modificar los parámetros *Xmx* y *XX*, y definir la cantidad de memoria necesaria. En la figura I.1 se presenta el ajuste realizado para este proyecto.



```
spoon.sh (~/Escritorio/LOD/data-integration5.4) - gedit
**
#
*****

if [ -z "$PENTAHO_DI_JAVA_OPTIONS" ]; then
    PENTAHO_DI_JAVA_OPTIONS="-Xmx8192m -XX:MaxPermSize=7168m"
fi

OPT="$OPT $PENTAHO_DI_JAVA_OPTIONS -Djava.library.path=$LIBPATH -
DKETTLE_HOME=$KETTLE_HOME -DKETTLE_REPOSITORY=$KETTLE_REPOSITORY -
DKETTLE_USER=$KETTLE_USER -DKETTLE_PASSWORD=$KETTLE_PASSWORD -
DKETTLE_PLUGIN_PACKAGES=$KETTLE_PLUGIN_PACKAGES -
DKETTLE_LOG_SIZE_LIMIT=$KETTLE_LOG_SIZE_LIMIT -DKETTLE_JNDI_ROOT=
$KETTLE_JNDI_ROOT"

# optional line for attaching a debugger
# OPT="$OPT -Xdebug -Xnoagent -Djava.compiler=NONE -
Xrunjdpw:transport=dt_socket,server=y,suspend=n,address=5005"

# *****|
# ** Run... **
# *****
"$PENTAHO_JAVA" $OPT -jar "$STARTUP" -lib $LIBPATH "${1+$@}"

sh Tab Width: 8 Ln 206, Col 18 INS
```

Figura I.1: Incremento de memoria a PDI

Nota: Tener en cuenta que la memoria asignada a PDI debe ser menor a la memoria con la que cuenta el ordenador, debido a que este deberá usarlo también para los procesos del sistema.

I.2.2. Almacenamiento de registros

La base de datos utilizada por el componente *Data Precatching* es configurada por defecto para almacenar solamente diez mil registros, si sabe que la cantidad de registros es mayor, deberá aumentar dicho valor en la transformación. Para configurar este valor se deberá dar doble click sobre el área de trabajo, con lo cual se abrirá una ventana de configuración de toda la transformación y en la pestaña *Micellaneous* se deberá especificar un valor cercano a la cantidad de registros con las que se trabajará.

En este proyecto se trabajó con alrededor de ocho millones de registros, para evitar inconvenientes en la configuración se especificó diez millones de registros para evitar cualquier tipo de inconveniente. Ver figura I.2.

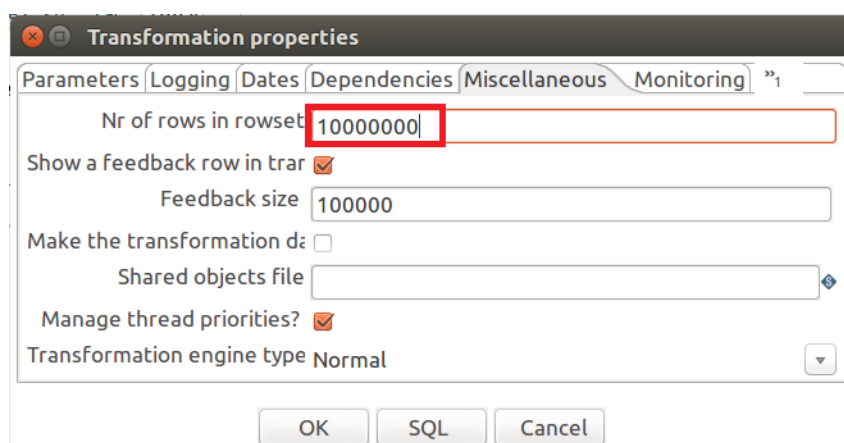


Figura I.2: Incremento de memoria a PDI

Nota: Si la cantidad de registros definida en la configuración es menor a la cantidad de registros leídos desde las fuentes de datos, el proceso podría nunca terminar.

I.2.3. Plugin Data Precatching

Se trata de un bug en el plugin, al usar este componente en la transformación no se incluye el tag que indica el nombre de la base de datos donde se almacenará los registros, por tanto, para corregir este problema se deberá abrir la transformación con un editor de texto y agregar el tag `<DBTABLE>OAIPMHDATA</DBTABLE>` manualmente, tal como se presenta en la figura I.3.

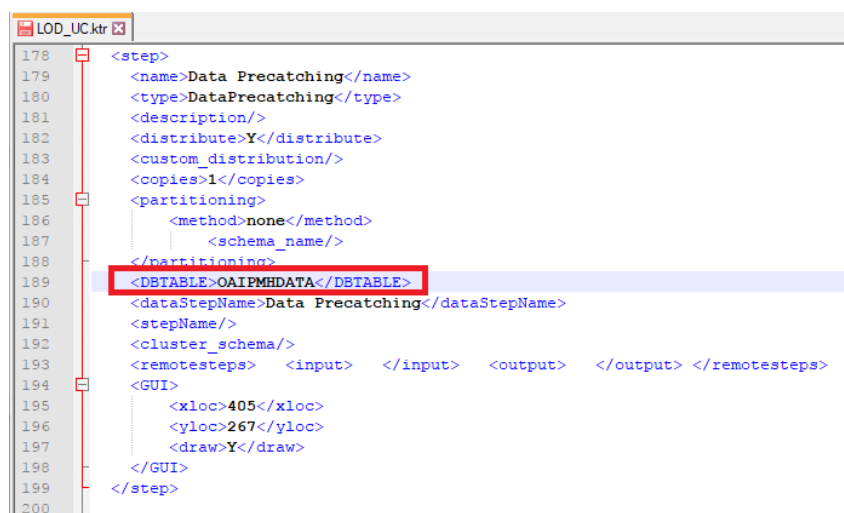


Figura I.3: Corrección de transformación

Nota: Si el tag no está incluido en la transformación los registros de las fuentes de datos no se almacenarán en la base de datos del plugin *Data Precatching* y por tanto no será posible la generación del RDF. Este problema podría depender de la versión de los plugins LOD con los que se trabaje.

I.2.4. Error en la base de datos interna

Es posible que la base de datos interna se vea afectada por alguna razón, si esto sucede se mostrará un mensaje similar al que se presenta en la figura I.4, para solucionar este problema se deberá eliminar los archivos *dblod.h2.db* y *dblod.trace.db*, estos archivos son creados automáticamente por el plugin *Ontology & Data Mapping* al abrir y aceptar. Dichos archivos se pueden encontrar en *home/nombre_de_usuario* para Linux y en *C:/Users/nombre_de_usuario* para Windows.

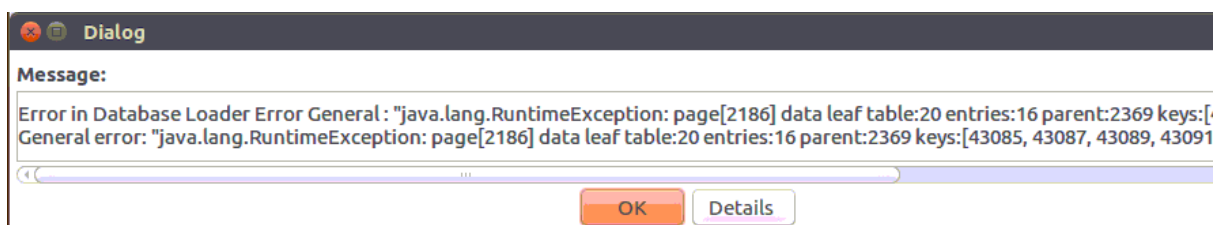


Figura I.4: Error en base de datos interna

I.2.5. Error en generación Elda

Al generar un proyecto *Elda* y desplegarlo es posible que algunos enlaces esten caídos, presentando el mensaje de error *HTTP ERROR 404*. Para solucionar este problema se deberá modificar manualmente las entidades mal generadas en el archivo *elda/webapps/ROOT/specs/recurso.ttl* del proyecto *Elda*.

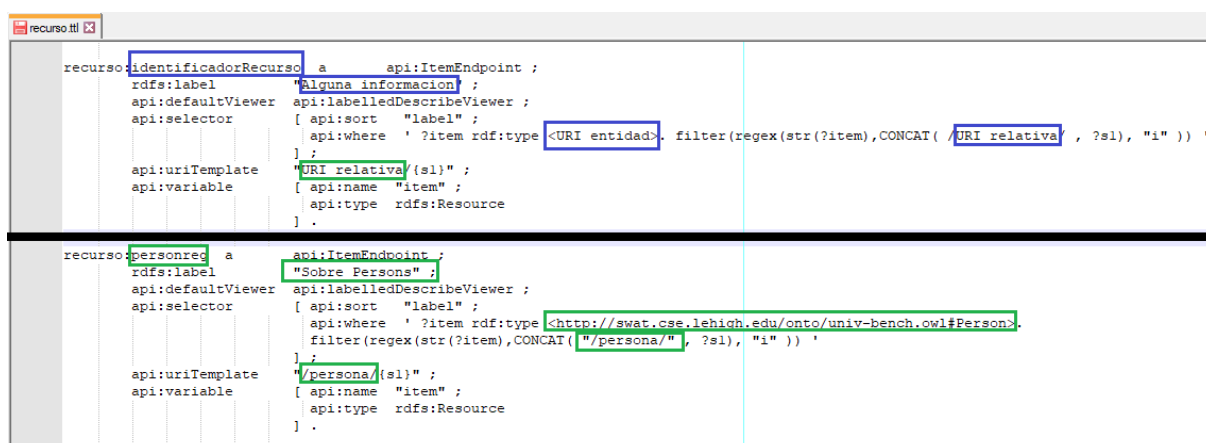


Figura I.5: Corrección de entidades - Elda

En la parte superior de la figura I.5 se presenta una plantilla para establecer un recurso que será presentado por el servicio *Elda*, tener en cuenta el *identificadorRecurso* puesto que este se deberá colocar en una sección del mismo archivo tal como se presenta en la figura I.6. Los siguientes valores a ajustar son: *label* para etiquetar a la entidad; *URI entidad* de la entidad definida en la ontología; y la *URI relativa* utilizada en el componente *Ontology & Data Mapping*.

Un ejemplo de esta configuración para la entidad *Person* se presenta en la parte inferior de la figura I.5 donde se ajustan todos los valores para tal recurso, posteriormente, la referencia al identificador *personreg* es colocado en *recurso:recurso* tal como se presenta en la figura I.6.

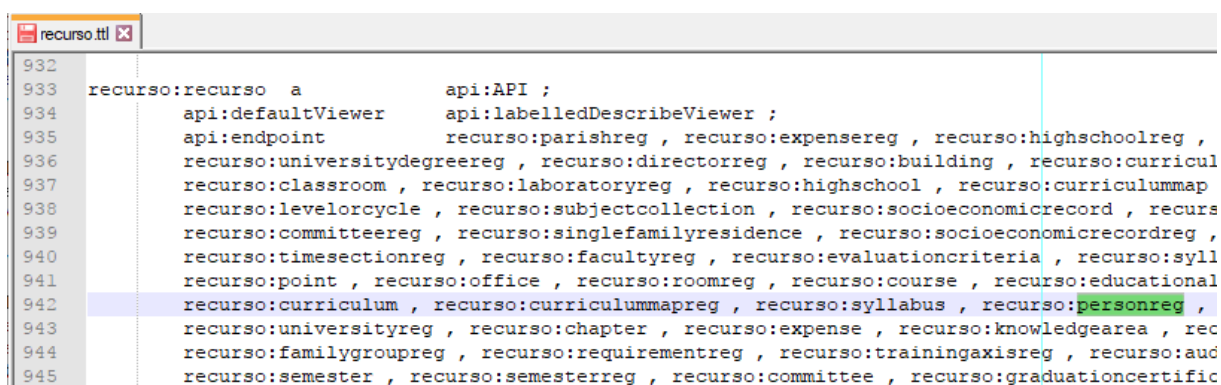


Figura I.6: Definición de referencias a recursos - Elda

Nota: Hacer esto en todos los recursos en los que se manifieste el error mencionado anteriormente.

I.3. Buscador semántico

En esta sección se tratará temas relacionados a los problemas enfrentados en el buscador semántico

I.3.1. Visualización gráfica

Si al elegir la opción ver grafo, el buscador no retorna ningún resultado se puede deber a que la base de datos del buscador contiene un valor nulo en sus registros, pasa solucionar este problema se realizará el siguiente proceso:

1. No detener el proceso correspondiente al buscador semántico.
2. Situararse en el directorio del buscador *sparql-fedquest/*, a través de la terminal de Linux.
3. Ejecutar el comando *meteor mongo* para iniciar la base de datos de mongo.
4. Para mostrar la información relacionada a un *endpoint* se ejecutará el comando *db.entities.find()*.
5. Copiar el valor correspondiente al campo *_id* del resultado anterior, y pegarlo en el comando del siguiente paso.
6. Para eliminar los valores nulos se ejecutará el comando *db.entities.update()* tal como se presenta en la figura I.7.

```
meteor:PRIMARY> db.entities.find()
{ "_id" : "tq5WhkjHKYJkchJd7", "endpoint" : "http://localhost:3030/academico/query", "graph" : "http://l
" : "Degree", "ent" : "type name", "dim" : "col-xs-6 col-sm-4 col-md-3" }, { "fullName" : "http://vivowe
", "dim" : "col-xs-6 col-sm-4 col-md-3" }, { "fullName" : "http://swat.cse.lehigh.edu/onto/univ-bench.owl:
col-sm-4 col-md-3" }, { "fullName" : "http://vivoweb.org/ontology/core#ResearchOrganization", "prefix" :
l-md-6" }, { "fullName" : "http://vivoweb.org/ontology/core#Committee", "prefix" : "", "name" : "Committ
purl.org/vocab/aiiso/schema#Faculty", "prefix" : "", "name" : "Faculty", "ent" : "type name BFO_0000050
v-bench.owl#University", "prefix" : "", "name" : "University", "ent" : "type BFO_0000051 isMainAuthority
meteor:PRIMARY> db.entities.update ({"_id" : "tq5WhkjHKYJkchJd7"}, {$pull :{'entities':null}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 0 })
```

Figura I.7: Corrección de vista gráfica

7. Iniciar nuevamente el buscador semántico.

Bibliografía

- [1] T. Berners-Lee, “Semantic web on xml,” 2000.
- [2] W. W. W. Consortium *et al.*, “Best practices for publishing linked data,” 2014.
- [3] M. C. Suárez-Figueroa, “Neon methodology for building ontology networks: specification, scheduling and reuse,” Ph.D. dissertation, Informatica, 2010.
- [4] M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, *Ontology engineering in a networked world*. Springer Science & Business Media, 2012.
- [5] G. Bustán, A. María, and D. F. J. González, “Principios y tecnologías linked data para la publicación de datos académicos de las diferentes carreras de la universidad nacional de loja.” B.S. thesis, 2016.
- [6] M. Jovanovik and D. Trajanov, “Consolidating drug data on a global scale using linked data,” *Journal of biomedical semantics*, vol. 8, no. 1, p. 3, 2017.
- [7] M. Poveda Villalon, “Metodología neon aplicada a la representación del contexto,” Ph.D. dissertation, Informatica, 2010.
- [8] K. Rodríguez Perojo and R. Ronda León, “Web semántica: un nuevo enfoque para la organización y recuperación de información en el web,” *Acimed*, vol. 13, no. 6, pp. 0–0, 2005.
- [9] A. Aquino. Evolución de la web. [Online]. Available: <http://jeuazarru.com/wp-content/uploads/2016/11/Evolucion-de-la-web.pdf>
- [10] J. J. Castaño and S. Jurado, *Marketing digital (Comercio electrónico)*. Editex, 2016.
- [11] I. Nafría, *Web .0: El usuario, el nuevo rey de Internet*. Gestión 2000, 2007.
- [12] J. Lopez. Evolución de la web. [Online]. Available: <https://es.slideshare.net/Nomada2070/definicion-y-caracteristicas-de-web-10-2030>

-
- [13] W. W. W. Consortium *et al.*, “Guía breve de web semántica,” *línea*. Available: <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>. [Último acceso: 20 Abril 2014], 2014.
- [14] M. Dürst and A. Freytag, “Unicode in xml and other markup languages,” Unicode Technical Report, Tech. Rep., 2002.
- [15] L. Masinter, T. Berners-Lee, and R. T. Fielding, “Uniform resource identifier (uri): Generic syntax,” 2005.
- [16] W. W. W. Consortium *et al.*, “Extensible markup language (xml) 1.1 (second edition),” 2006.
- [17] D. C. Fallside and P. Walmsley, “Xml schema part 0: primer second edition,” *W3C recommendation*, vol. 16, 2004.
- [18] D. Beckett and B. McBride, “Rdf/xml syntax specification (revised),” *W3C recommendation*, vol. 10, no. 2.3, 2004.
- [19] D. Brickley and R. Guha. Rdf schema 1.1. [Online]. Available: <https://www.w3.org/TR/rdf-schema/>
- [20] D. Hurtado. Introducción a la web semántica: Realidades y perspectivas. [Online]. Available: <http://jeuazaru.com/wp-content/uploads/2016/11/Evolucion-de-la-web.pdf>
- [21] A. L. Tello, “Ontologías en la web semántica,” *España: Universidad De Extremadura*, 2001.
- [22] J. Heflin, “Owl web ontology language-use cases and requirements,” *W3C Recommendation*, vol. 10, p. 12, 2004.
- [23] La ontología y la web semántica: Recomendaciones de la w3c. [Online]. Available: <http://www.bibliopos.es/Biblion-A2-Bibliografia-Documentacion/18ontologia-Web-Semantica.pdf>
- [24] N. Valencia, “Creación de una ontología para el dominio de e-gobierno en México,” 2011.
- [25] A. C. Muñoz García, B. Sandia Saldivia, and G. Páez Monzón, “Un modelo ontológico para el aprendizaje colaborativo en la educación interactiva a distancia,” *Educere*, vol. 18, no. 61, 2014.
- [26] W. W. W. Consortium *et al.*, “Guía breve de linked data,” *línea*. Available: <http://www.w3c.es/Divulgacion/GuiasBreves/LinkedData>, 2014.

-
- [27] C. Laufer, “Guía de la web semántica,” 2015.
- [28] R. C. Fernández, “Representación del conocimiento. web semántica,” *Universidad Carlos III de Madrid*, 2008.
- [29] N. Zeas Orellana, “Desarrollo de una plataforma web para una red de monitorización de calidad del aire,” Master’s thesis, Universitat Politècnica de Catalunya, 2015.
- [30] W. W. W. Consortium *et al.*, “Sparql lenguaje de consulta para rdf,” 2008.
- [31] G. Antoniou and F. Van Harmelen, “Web ontology language: Owl,” in *Handbook on ontologies*. Springer, 2004, pp. 67–92.
- [32] J. A. P. Sánchez, *Tecnologías de la web semántica*. Editorial UOC, 2011.
- [33] B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue, C. Lutz *et al.*, “Owl 2 web ontology language profiles,” *W3C recommendation*, vol. 27, p. 61, 2012.
- [34] B. Hyland and B. Villazón Terrazas, “Linked data cookbook,” 2012.
- [35] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, “Methodological guidelines for publishing government linked data,” *Linking government data*, pp. 27–49, 2011.
- [36] M. C. Suárez-Figueroa, A. Gómez-Pérez, and B. Villazón-Terrazas, “How to write and use the ontology requirements specification document,” in *OTM Confederated International Conferences. On the Move to Meaningful Internet Systems*. Springer, 2009, pp. 966–982.
- [37] M. B. Mora Arciniega and V. Segarra Faggioni, “Modelo ontológico para la representación de datos académicos y su publicación con tecnología semántica,” *Opción*, vol. 32, no. 10, 2016.
- [38] B. Szásza, R. Fleinerb, and A. Micsikc, “Oloud—an ontology for linked open university data.”
- [39] Y. Rosell León, J. A. Senso Ruiz, and A. A. Leiva Mederos, “Diseño de una ontología para la gestión de datos heterogéneos en universidades: marco metodológico,” *Revista Cubana de Información en Ciencias de la Salud*, vol. 27, no. 4, pp. 545–567, 2016.
- [40] L. Zemmouchi-Ghomari and A. Ghomari, “Process of building reference ontology for higher education,” in *Proceedings of the World Congress on Engineering*, vol. 3, 2013, pp. 1595–1600.
- [41] U. de Cuenca. Lod-gf: An integral open data generation framework. [Online]. Available: <https://github.com/ucuenca/lodplatform>
-



- [42] J. Rengel Quiroga, “Desarrollo de ontología de títulos de grado,” 2015.
- [43] R. Jarrín, S. Rodrigo, and J. J. Sáenz Peñafiel, “Modelamiento del perfil de un usuario usando tecnologías semánticas,” B.S. thesis, 2014.
- [44] S. Das, S. Sundara, and R. Cyganiak, “R2rml: Rdb to rdf mapping language,” 2012.